

Faculdade de Engenharia da Universidade do Porto



**Previsão de procura intermitente no setor do
retalho de distribuição alimentar**

Gonçalo Fernando de Pinho e Silva de Almeida Falcão

VERSÃO FINAL

Dissertação realizada no âmbito do
Mestrado Integrado em Engenharia Eletrotécnica e de Computadores
Ramo Automação

Orientador: Prof. Dr. Américo Lopes de Azevedo
Co-orientador: Prof. Dra. Patrícia Alexandra Gregório Ramos

22 de junho de 2017

© Gonalo de Pinho Falco, 2017

Resumo

O trabalho aqui desenvolvido é resultado de um acordo realizado entre o INESC TEC e o Grupo Jerónimo Martins, que disponibilizou um vasto conjunto de dados relativo à procura registada em 412 lojas deste retalhista sobre a sua marca Pingo Doce, com o objetivo de possibilitar a este centro de investigação a análise e estudo aprofundado da procura de carácter intermitente.

Dado o crescente interesse no estudo da intermitência na previsão, característica que se mostra dominante no conjunto de dados disponibilizado, e em resultado da existência de enormes lacunas na investigação desta área de conhecimento, que se encontra ainda nos princípios do seu desenvolvimento, o trabalho aqui desenvolvido e os resultados nele obtidos constituem um valioso contributo.

Nesta análise foi levado a cabo um pré-processamento do conjunto de dados por forma a que estes pudessem apresentar a estrutura viável à aplicação dos métodos de previsão adequados para séries de dados intermitentes cujo levantamento exaustivo, foi, entretanto, efetuado.

Com recurso à linguagem R, foi desenvolvido um procedimento experimental baseado em *rolling*, que permitiu avaliar a eficiência de previsão de todos esses métodos com recurso a medidas de erro especialmente concebidas para séries com elevado grau de intermitência.

Os resultados obtidos permitiram mostrar que a implementação dos métodos de previsão para procura intermitente no seio de uma *framework* de modelos de espaço de estados permite obter uma melhor performance de previsão.

Abstract

The work developed here is the result of an agreement for INESC TEC and the Jerónimo Martins Group, which made available a vast data set in relation to the demand in 412 stores of this retailer on its Pingo Doce brand, in order to enable this research center to analyze and deepen the study for intermittent demand.

Given the growing interest in the study of intermittent prediction, which is a dominant feature in the data set available, and as a result of the existence of huge gaps in the research of this area of knowledge, which is still at the beginning of its development, the work here and the results obtained from it are a valuable contribution.

In this analysis a preprocessing of the data set was carried out so that they could present the viable structure to the application of the appropriate forecasting methods for intermittent data series, whose exhaustive survey has been made in the meantime. With the use of the R language, an experimental procedure based on rooling was developed, which allowed to evaluate the forecasting efficiency of all these methods with the use of error measures specially designed for series with a high degree of intermittence. The results obtained showed that the implementation of forecasting methods for intermittent search within a framework of state space models allows to obtain a better prediction performance.

Agradecimentos

A realização desta dissertação de mestrado culmina do trabalho dos passados meses e do empenho de vários anos até ao presente momento. Sendo, como tal, importante demarcar os papéis que diversas pessoas tiveram ao longo desta jornada.

À instituição que me acolheu, INESC TEC, representada especificamente pela Sra. Professora Doutora Patrícia Alexandra Gregório Ramos a quem agradeço todo o apoio, total disponibilidade, ajuda, incentivo e orientação que tão amavelmente me prestou no decorrer destes meses e que me possibilitou chegar com sucesso ao dia de hoje.

Ao meu orientador, Professor Doutor Américo Lopes de Azevedo, agradeço a disponibilidade, orientação e ação em momentos fundamentais no desenrolar deste trabalho de dissertação.

A todos, os poucos, amigos que, mais ou menos presentes, acompanharam-me ao longo destes vários anos na FEUP, mas, acima de tudo, ao meu grande amigo Tiago Leite Constante, companheiro em todos os passos desta demanda, tanto da FEUP como deste trabalho em específico, esperando que continuemos sempre presentes na vida um do outro, e que a vida de cada um seja uma longa viagem, e que continue assim camarada: difícil, mas interessante.

Com elevado um peso nesta jornada decorrida, queria agradecer a toda a minha família todo o seu amor, carinho e investimento ao longo destes diversos anos por forma a permitir-me atingir o sucesso académico. Agradeço tanto aos meus pais, por todo o seu carinho e incentivo, assim como ao meu padrinho, Eng. Francisco Queirós de Castro e a minha tia Maria Manuela Martins Castro, por nunca terem deixado de apostar em mim ao longo destes anos.

Por fim, deixo uma nota muito especial aos meus companheiros de vida, os meus dois irmãos, o Prof. Doutor Nuno de Pinho Falcão e o Engenheiro Luís de Pinho Falcão, que num percurso de vida tão atribulado nunca em momento algum deixaram de estar ao meu lado, desde o primeiro dia das nossas vidas. Não só esta dissertação, como todo o meu percurso académico e os feitos daí decorrentes, são-vos dedicados.

Gonçalo de Pinho Falcão

*“Educating the mind
without educating the heart
is no education at all.”*

Aristotle

Índice

Resumo	iii
Abstract.....	v
Agradecimentos	vii
Índice	x
Lista de figuras	xii
Lista de tabelas	xiii
Abreviaturas e Símbolos	xiv
Capítulo 1	1
Introdução	1
1.1 Importância da previsão do consumo	1
1.2 Enquadramento do projeto e motivação	4
1.3 Objetivos da dissertação	6
1.4 Estrutura do documento	6
Capítulo 2	9
Revisão bibliográfica	9
2.1. Métodos de previsão de procura intermitente.....	9
2.1.1. Método Naïve	11
2.1.2. Simple Exponential Smoothing	12
2.1.3. Croston	15
2.1.4. Aproximação de Syntetos-Boylan	21
2.1.5. Teunter, Syntetos e Babai.....	26
2.1.6. Intermittent Multiple Aggregation Prediction Algorithm	31
2.2. Modelos de espaço de estados.....	40
2.2.1. Modelo de espaço de estados puramente aditivo para intermitência	43
2.2.2. Modelo espaço de estados puramente multiplicativo para intermitência	44
2.2.3. Croston	45
2.2.4. Teunter, Syntetos e Babai.....	45
2.2.5. Fixed e Auto	46
Capítulo 3	49
Caso de Estudo	49
3.1. Caso Jerónimo Martins	49
3.2. Conjunto de dados.....	49
3.3. Procedimento experimental	52
3.3.1. Processo de rolling	52
3.3.2. Métodos e pacotes de software	54
3.3.3. Medidas de avaliação de previsão	55
3.4. Resultados	59
Capítulo 4	61
Conclusões e Trabalho Futuro	61

Referências	64
Anexos	68
A.1 - Combinações da sazonalidade e da tendência	68
A.2 - Classificação proposta por Pegels(1969)	68
A.3 - Métodos e valores iniciais	69
A.4 - Formulas de cálculo recursivas	70
A.5.1 - Equações de espaço de estados para os modelos na estrutura ETS - Erros Aditivos.....	71
A.5.2 - Equações de espaço de estados para os modelos na estrutura ETS - Erros Multiplicativos	72
A.6 - Resultados obtidos com MsE, MdsE, MsAE e MdsAE	73
A.7 - Resultados obtidos com MsSE, MdsSE, MAsE e MdAsE	74
A.8 - Resultados obtidos com MsPIS e MsAPIS.....	75

Lista de figuras

Figura 1. Impacto do Stock-out no consumidor (Fonte: HarvStudy).....	2
Figura 2. Impacto da previsão no planeamento nos níveis de inventário (Fonte: IBM Ilog)....	2
Figura 3. Alterações na procura independentes de alterações do preço.....	3
Figura 4. Jerónimo Martins no mundo (Fonte: Jerónimo Martins).....	4
Figura 5. Percurso de mercado do grupo Jerónimo Martins entre Dez '10 e atualidade (Fonte:Plus500).....	5
Figura 6 - Exemplo de previsão por recurso ao <i>Naïve</i> , <i>Naïve</i> sazonal e média (Hyndman et al,2013).....	11
Figura 7 - Exemplo da decomposição de Croston (Petropoulos, 2016).....	16
Figura 8 - Agregação temporal sem sobreposição com 12 níveis.....	32
Figura 9 - Previsão com ADIDA (Nikolopoulos et al, 2011).....	34
Figura 10 - MASE dos níveis de agregação para os vários métodos de previsão/desagregação (Nikolopoulos et al.,2011).....	35
Figura 11 - Esquema de categorização SBC (Kostenko, 2006).....	38
Figura 12 - Espaço de parâmetros para os quais um método tem melhor desempenho que o outro.....	39
Figura 13 - Processo de rolling com horizonte de previsão a um passo.....	53
Figura 14 - Processo de rolling com horizonte de previsão de quatro períodos.....	54

Lista de tabelas

Tabela 1 - Exemplo de aplicação do Simple Exponential Smoothing.....	14
Tabela 2 - Determinação da componente \hat{z}_t do método de Croston.....	18
Tabela 3 - Determinação da componente \hat{p}_t do método de Croston.....	19
Tabela 4 - Determinação da previsão e resíduo do método de Croston.....	21
Tabela 5 - Determinação da componente \hat{z}_t do SBA.....	25
Tabela 6 - Determinação da componente \hat{p}_t , previsão e erro do SBA.....	26
Tabela 7 - Determinação da componente \hat{z}_t do TSB.....	29
Tabela 8 - Exemplo de aplicação do TSB, Determinação de dt , previsão e erro.....	30
Tabela 9 - Exemplo da variação do grau de intermitência.....	37
Tabela 10 - Número de SKUs por área de atividade.....	51
Tabela 11 - Estatística descritiva do conjunto de dados.....	52
Tabela 12 - Exemplo de aplicação PIS.....	58
Tabela 13 - <i>Rank</i> médio de cada método para os diferentes horizontes e <i>Rank</i> global.....	59

Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

ADIDA	Aggregate-Disaggregate Intermittent Demand Approach
ASE	Absolute Scaled Error
B2B	Business-to-Business
B2C	Business-to-consumer
CESE	Centro de Engenharia de Sistemas Empresariais
CRO	Croston
DEEC	Departamento de Engenharia Eletrotécnica e de Computadores
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortization
EQW	Equal Weights
FEUP	Faculdade de Engenharia da Universidade do Porto
FMI	Fast-Moving Items
GMRMSE	Geometric Mean Relative Mean Square Error
iMAPA	intermittent Multiple Aggregation Prediction Algorithm
INESC TEC	Instituto de Engenharia de Sistemas e Computadores - Tecnologia e Ciência
JM	Jerónimo Martins SGPS, S.A.
MA	Moving Average
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
MdASE	Median Absolute Scaled Error
MsE	Mean Scaled Error
MSE	Mean Squared Error
PIS	Periods In Stock
sAE	scaled Absolute Error
SBA	Syntetos-Boylan Approximation
SES	Simple Exponential Smoothing
SKU	Stock Keeping Unit
sMAE	scaled Mean Absolute Error
sMdAE	scaled Median Absolute Error

sMdSE	scaled Median Squared Error
SMI	Slow-Moving Items
sMSE	scaled Mean Squared Error
sSE	scaled Squared Error
TSB	Teunter, Syntetos and Babai

Lista de símbolos

α	Parâmetro de alisamento do método SES
----------	---------------------------------------

Capítulo 1

Introdução

A presente dissertação em ambiente empresarial resultou de uma proposta efetuada pelo Instituto de Engenharia de Sistemas e Computação (INESC TEC) à Faculdade de Engenharia, tendo sido a mesma orientada e desenvolvida no Centro de Engenharia de Sistemas Empresariais (CESE).

Assim, no presente capítulo, é realizada uma breve análise e contextualização acerca da temática em foco. Em primeiro lugar, é feita uma análise da importância da previsão do consumo na vida quotidiana. De seguida é feito um enquadramento do tema em análise e são identificados os objetivos deste projeto. Por fim é especificada a estrutura do documento em questão.

1.1 Importância da previsão do consumo

A previsão é o processo de gerar um prognóstico acerca de um evento ou atividade que irá decorrer num momento futuro tendo por base dados passados. Podemos definir um sistema de previsão como “um conjunto de técnicas ou ferramentas necessárias para a análise de um histórico de dados, seleção de uma estrutura de modelação apropriada, desenvolvimento de previsões e monitorização e ajuste das mesmas” (Business Dictionary).

Tendo em conta que o foco deste trabalho de dissertação são os sistemas de previsão aplicados ao consumo, compreendemos rapidamente que um sistema destes que seja eficiente e eficaz pode ser visto como uma grande vantagem competitiva para uma empresa e para a posição que esta ocupa na secção de mercado na qual está inserida.

Com os padrões de consumo que se verificam nos dias de hoje e o ritmo a que este consumo ocorre é cada vez de maior importância para as empresas, e para o sucesso das suas operações, conseguirem efetuar previsões o mais próximo possível da procura que efetivamente é registada, por forma a que lhes seja possibilitado ajustar os seus mecanismos de produção e logística em tempo real e assim satisfazer com sucesso a procura dos seus produtos e serviços.

Esta vantagem visa dotar o utilizador de ferramentas que lhe permitam evitar situações que lhe sejam indesejáveis tais como, no caso de um retalhista, a situação de *stockout*, resultante da disponibilização de uma quantidade de SKUs inferior à procura que

posteriormente é registrada, o que prejudica a imagem que o cliente tem do retalhista. Segundo um estudo realizado por Corsten et al. (2004), apresentado na Harvard Business Review (ver Figura 1), 7% a 25% dos consumidores continuam a comprar, mas sem adquirir um substituto para o produto em falta, e cerca de 21% a 43% dos clientes preferem recorrer à concorrência, acarretando perdas que rondam os 4% nas vendas do retalhista. Ou ainda, nas situações de excesso de stock, que acarretam custos de operação superiores e consequente diminuição do lucro da atividade, elevados gastos com armazenamento e transporte, possibilidade de deterioração e obsolescência dos produtos ou mudanças nos padrões de procura do mercado.

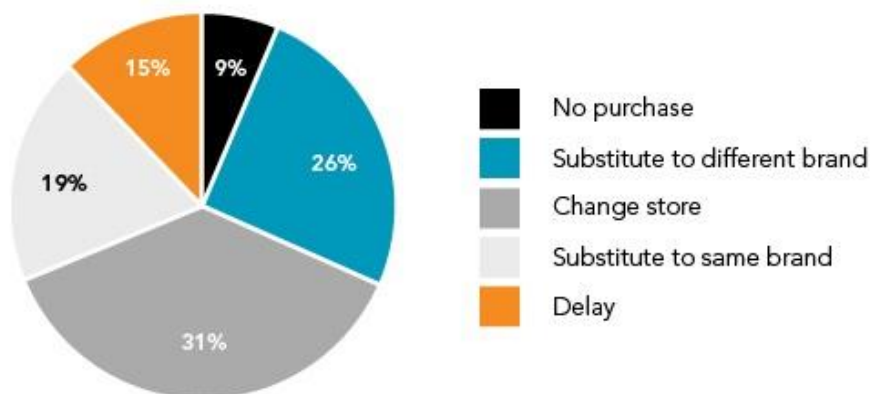


Figura 1. Impacto do *stock-out* no consumidor (Fonte: HarvStudy).

Torna-se assim simples compreender o extremo relevo que o estudo e desenvolvimento destas competências de previsão têm, não só para a indústria como também a sua aplicabilidade em muitas outras áreas, no nosso dia-a-dia, no mundo que nos rodeia e o impacto positivo que esta capacidade pode conferir a uma empresa (ver Figura 2).

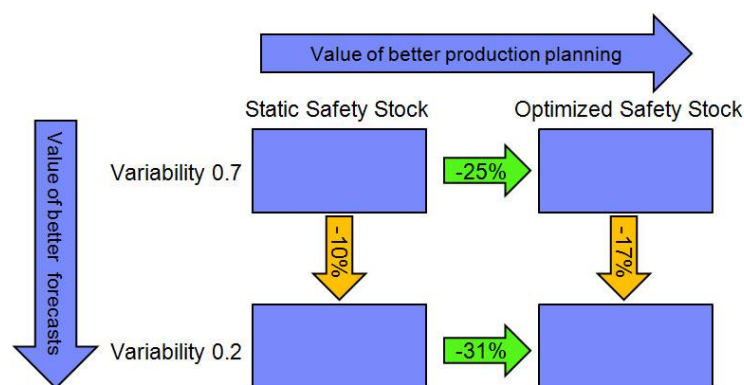


Figura 2. Impacto da previsão no planeamento dos níveis de inventário (Fonte: IBM Ilog).

A figura 2 representa o impacto que o planeamento da produção e a performance da previsão têm para os níveis de inventário. Podemos concluir com a imagem que quanto

melhor planeada for a produção maior é a otimização do *safety stock*, e quanto melhores forem as previsões dos analistas menor é a variabilidade registada.

Na situação ideal, ou seja, com o melhor planeamento de produção e as melhores previsões de consumo é possível registar melhorias de cerca de 31% nos níveis de inventário e de cerca de 17% na variabilidade dos SKUs.

De uma forma simples, a previsão da procura baseia-se no estudo de um histórico de dados de consumo de um determinado produto ou serviço com o objetivo de apreender comportamentos ou padrões presentes nesse conjunto de dados que possibilitem a aplicação e desenvolvimento de modelos para extrapolação do comportamento futuro da procura desses mesmos serviços ou produtos.

É importante também entender que a aplicabilidade destes métodos e interpretação dos resultados daí provenientes deverão ser analisados num contexto mais alargado tendo em conta os diversos fatores que podem influenciar a procura de um determinado produto ou serviço. Quando refletimos sobre quais são esses fatores um dos mais imediatos será o preço que o produto ou serviço tem, contudo existem muitos outros fatores com elevado relevo. Fatores como a população alvo, as expectativas do consumidor, a capacidade de compra do público ou a sazonalidade e as modas têm um grande impacto na forma como a procura é registada, pelo que têm de ser tidos em conta quando se realiza um estudo deste tipo (ver Figura 3).

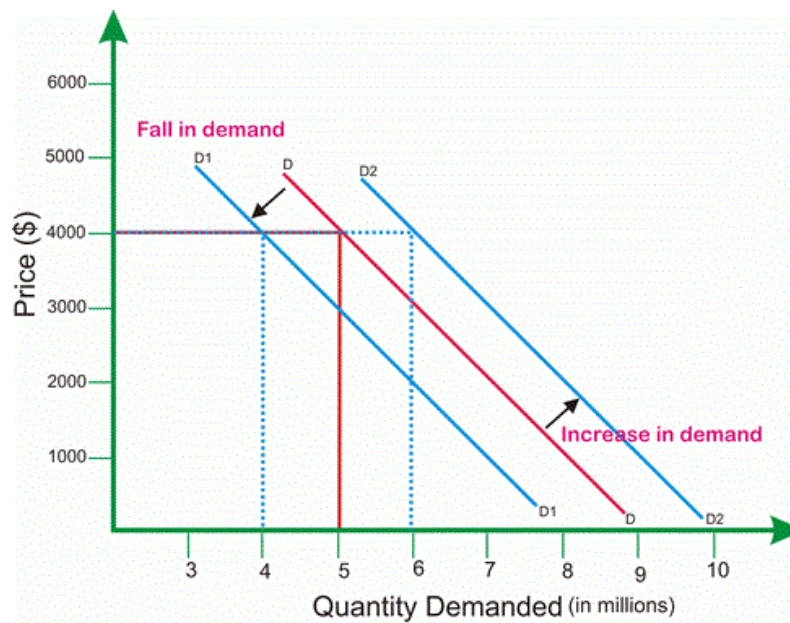


Figura 3. Alterações na procura independentes de alterações do preço.

1.2 Enquadramento do projeto e motivação

A presente dissertação surge no contexto de um trabalho iniciado pela Professora Patrícia Ramos no CESE no âmbito de um projeto que envolve o estudo e análise de dados de consumo compreendidos entre 3 de janeiro de 2012 e 27 de abril de 2015 pertencentes à empresa Jerónimo Martins SGPS, S.A., mais especificamente às vendas de lojas Pingo Doce.

O grupo Jerónimo Martins SGPS, S.A. é um grupo que teve a sua origem em 1792, numa loja aberta no Chiado, em Lisboa, por um Galego que mais tarde, em 1920, é comprada por dois empresários do Porto. Em 1978 este grupo entra no negócio da distribuição alimentar moderna com a abertura das suas lojas Pingo Doce. Em 1982 o grupo celebra uma *joint venture* com o grupo Belga Delhaize, permitindo-lhe assim um franco crescimento na área do retalho alimentar com partilha de *know how* entre ambos. Em 1992 a Jerónimo Martins cria uma parceria estratégica com a empresa retalhista Holandesa Ahold, que em julho de 2016 passa, por fusão, a ser Ahold Delhaize, com o objetivo de potenciar ainda mais o crescimento e expansão da sua marca Pingo Doce.

Este grupo é líder do segmento do retalho alimentar em Portugal, tendo sido o primeiro com loja on-line e visto o seu número de lojas triplicar entre 2002 e 2009, contando na atualidade com mais de 400 lojas em Portugal.

O grupo possui também um investimento com elevado sucesso na secção de retalho alimentar na Polónia, com a marca Biedronka a ser a maior cadeia de supermercados nesse país com mais de 2000 lojas, a marca Recheio, investimento que realizaram na área grossista alimentar e mais recentemente, em 2012, expandiram para a Colômbia investindo, também no retalho alimentar utilizando a insígnia Ara. À data, este grupo possui cerca de 3398 lojas em todo o mundo com um valor de vendas de 14.453 milhões de euros, um EBITDA de 862 milhões de euros e possuindo no total 28 centros de distribuição entre Portugal, Polónia e a Colômbia.



Figura 4. Jerónimo Martins no mundo (Fonte: sítio web do Grupo Jerónimo Martins).

Analisando o percurso e dimensão que este grupo, detentor da marca Pingo Doce, adquiriu ao longo dos anos consegue-se entender a elevada quantidade de bens de consumo que um retalhista desta dimensão tem de gerir, e a quantidade de variáveis que influenciam a procura pelos mesmos, tornando-se crucial o estudo e previsão do comportamento da procura de que são alvo esses múltiplos produtos (ver Figura 5).

Quando pensamos em comercialização de produtos, podemos dividir a procura destes bens de consumo essencialmente em duas categorias: *fast-moving items* (FMI) e *slow-moving items* (SMI). Os FMI são os produtos com elevado volume de vendas, normalmente de preço relativamente baixo, e por serem de comercialização “rápida” são também conhecidos por “*fast moving goods*”. Enquanto que os SMI são produtos que apresentam baixas taxas de procura em comparação com os primeiros. De uma forma prática, estes produtos caracterizam-se por apresentarem zero unidades vendidas nos seus registos ao longo do tempo criando o que se chama uma intermitência na procura, sendo assim considerados de comercialização “lenta” pelo que são também conhecidos por “*slow moving goods*”.



Figura 5. Percurso de mercado do Grupo Jerónimo Martins entre Dez '10 e a atualidade (Fonte: Plus500)

O foco desta dissertação prende-se com a análise e previsão da procura dos SMI do conjunto de dados referido, ou seja, categorias de produtos que apresentam percentagem de procura inferior a 100% no conjunto de dados da série temporal em análise, ou seja, em períodos temporais sucessivos com igual periodicidade.

A intermitência no consumo é um tópico com maior relevo do que aquele que muitas vezes lhe é reconhecido e uma área científica em desenvolvimento e alvo de profunda investigação uma vez que tem ainda amplo terreno para progredir e evoluir.

Apesar de, numa primeira fase, os produtos com procura intermitente poderem parecer bens de consumo que não apresentam grande interesse enquanto foco de análise para o retalhista, uma vez que não apresentam um elevado volume de vendas, faço denotar a sua

elevada importância e o enorme relevo que esta parcela de *Stock Keeping Units* (SKUs) tem para a atividade de uma empresa, uma vez que estes representam, em média, o maior volume de unidades constantes da atividade dos retalhistas, podendo no seu conjunto superar uma percentagem de 60% do valor total de SKUs segundo o estudo de Banker(2013). Para compreendermos, num exemplo mais concreto, a importância dos SKUs de procura intermitente para a área do retalho, no conjunto de dados que é analisado neste trabalho, o volume de SKUs que apresentam procura intermitente supera 80% do total de SKUs que este retalhista comercializa. Se refletirmos no relevo que estes produtos têm para o volume de negócios de uma empresa compreendemos a franca necessidade do seu estudo e aprofundamento em termos de análise da procura bem como o desenvolvimento de metodologias que permitam obter previsões cada vez mais corretas que favoreçam o desenvolvimento da atividade dos retalhistas.

1.3 Objetivos da dissertação

Fundamentalmente, o objeto de estudo desta dissertação é a previsão da procura de *slow-moving* de itens do setor do retalho de distribuição alimentar, pretendendo-se realizar uma análise univariada do conjunto de dados cedido que permita prever o consumo dos SKUs desse retalhista para 52 semanas de vendas. Nesta análise avaliar-se-á a *performance* de previsão dos métodos e modelos estudados recorrendo a medidas de erro adequadas para séries temporais intermitentes. Esta análise é levada a cabo recorrendo a pacotes específicos para o efeito e existentes no *software R*. O objetivo é, após a obtenção dos resultados, concluir qual o método ou modelo mais adequado para a previsão de procura intermitente e aconselhar os retalhistas para a sua utilização. Deve referir-se que, do conhecimento resultante da extensiva procura na literatura, compreendeu-se não existir qualquer estudo do género do que é levado a cabo neste trabalho.

Resumindo, a sequência de etapas a seguir para o desenvolvimento da dissertação são:

- Análise, entendimento e aprofundamento do estado da arte associado com o tema em foco;
- Estudo e compreensão da estrutura dos dados em análise;
- Determinação dos métodos e modelos de previsão que melhor se adequam aos dados em questão, bem como a seleção das melhores práticas de avaliação empírica;
- Desenvolvimento computacional de uma metodologia de treino, previsão e avaliação recorrendo a *software* adequado ao efeito;
- Interpretação dos resultados obtidos por forma a compreender quais os métodos ou modelos que melhor se adequam ao caso em estudo.

1.4 Estrutura do documento

A dissertação aqui apresentada encontra-se estruturada em quatro capítulos.

Estes estão divididos para que do segundo capítulo conste o estado da arte, onde é aprofundado o estado de desenvolvimento em que se encontra a área científica em foco até ao momento de realização desta dissertação, bem como exploração mais pormenorizada do

funcionamento dos métodos de previsão que foram utilizados na realização deste trabalho, o seu suporte teórico e demonstração do seu funcionamento prático por forma a permitir um entendimento mais explícito e concreto do trabalho desenvolvido na análise do conjunto de dados em estudo bem como do grau de dificuldade associado com a análise de conjuntos de dados intermitentes e métodos que lhe estão associados.

Do terceiro capítulo consta uma análise do caso de estudo alvo da dissertação e de todo o processo experimental levado a cabo sobre esses dados, ou seja, contextualização do conjunto de dados disponíveis para estudo, introdução e explicação do programa desenvolvido em linguagem de programação R e especificidades metodológicas dos pacotes utilizados no seu desenvolvimento. Realiza-se também uma descrição das medidas de erro utilizadas para aferir a eficiência dos métodos bem como uma explicação do seu funcionamento e para terminar é realizada uma discussão dos resultados obtidos

Por fim, no quarto capítulo são elencadas as conclusões obtidas com a realização deste trabalho e dadas sugestões de possíveis melhorias, ou continuidade do aprofundamento dos assuntos abordados, decorrentes deste estudo.

A estrutura de documento adotada segue as regras que são estipuladas pelo Departamento de Engenharia Eletrotécnica e de Computadores, da Faculdade de Engenharia da Universidade do Porto, utilizando para o efeito o modelo de documento que é disponibilizado por esta instituição e seguindo as orientações que estão estabelecidas e indicadas no mesmo ficheiro, assim como recomendações provenientes dos respetivos orientadores da dissertação.

Capítulo 2

Revisão bibliográfica

O presente capítulo aborda todos os pontos que são importantes referir e essenciais ao entendimento da temática em discussão para a realização do estudo do estado da arte.

Assim, é efetuada, em primeiro lugar, uma introdução à procura intermitente e posteriormente são apresentados e aprofundados os diversos métodos utilizados para fazer previsão desse tipo de consumo, o seu funcionamento e evolução até ao momento presente.

2.1. Métodos de previsão de procura intermitente

A previsão é uma tarefa de grande importância, de maior ou menor complexidade consoante o que se pretende prever, e uma ferramenta que permite uma ajuda eficiente e eficaz à gestão em atividades que podem ser tão diferentes como prever o consumo de um produto numa loja para determinar o volume de *stock* necessário, prever o crescimento de atividade de determinada empresa por forma a compreender se é necessária a expansão das suas instalações ou até mesmo prever o crescimento da população mundial.

A previsão pode ser de dois tipos: qualitativa ou quantitativa. A primeira é maioritariamente aplicada em casos de inexistência ou insuficiência de dados históricos que permitam recorrer a métodos de previsão, enquanto que a segunda é utilizada quando existe um conjunto de dados históricos relevantes e é possível assumir que esses padrões serão repetidos em situações de consumo futuro.

Assim, encontramos, de uma forma natural, a aplicação destas técnicas de previsão associadas à atividade de empresas e análise das suas cadeias de abastecimento com foco na atividade *business-to-consumer* (B2C), porém esta perspetiva de utilização de ferramentas de previsão tem, nos últimos anos, sido também alvo de reflexão e recurso por parte daqueles que desenvolvem atividade *business-to-business* (B2B) conforme foi discutido e defendido no evento “*Sales Operations Planning Innovation*” em Boston e aprofundado por Banker (2013). As metodologias de previsão, na sua globalidade, contam já com um percurso extenso de investigação e desenvolvimento.

A intermitência no consumo apresenta-se como uma dificuldade no processo de previsão sendo, por isso, uma área menos conhecida e ainda em franco desenvolvimento com abordagens muito recentes, sendo que, as que foram desenvolvidas até à atualidade carecem

de aprofundamento em termos de investigação. Este fator demonstrou-se uma dificuldade considerável no desenvolvimento desta dissertação devido à deficiente disseminação de metodologias de análise aplicáveis, uma vez que sendo uma área menos estudada na previsão tem ainda reduzido suporte científico e dentro do conjunto de métodos disponíveis estes reclamam claramente de um desenvolvimento e aperfeiçoamento.

As especificidades que a intermitência tem em termos de comportamento na previsão levam a que se tente evitar as análises deste tipo de produtos. Tanto assim é que, alguns dos métodos aplicados neste tipo de casos recorrem a processos de redução da incidência de zeros no conjunto de dados e até idealmente a sua eliminação por completo afastando assim a sua análise do domínio da intermitência, o que simplifica a aplicação de métodos de previsão sofisticados para os quais existe um *background* de pesquisa bastante desenvolvido.

Por forma a compreendermos o efeito da intermitência na previsão da procura, é primeiramente necessário compreender que os métodos que são utilizados para prever a procura de séries¹ não intermitentes, sem zeros no seu conjunto de dados, não têm um comportamento aceitável quando aplicados a séries intermitentes originando desempenhos insatisfatórios. Assim, o conjunto de métodos e modelos que foram ponderados e utilizados no conjunto de dados que nos foi cedido pelo Grupo Jerónimo Martins são específicos, e foram desenvolvidos para serem aplicados em séries de dados intermitentes.

Este conjunto de dados fornecidos pelo Grupo Jerónimo Martins ao CESE contempla dados diários referentes às vendas de SKUs registadas em 412 lojas desse grupo. Estes dados foram fornecidos em bruto, no mesmo formato com que são geridos internamente pela empresa, tendo sido posteriormente desenvolvido um trabalho de tratamento dos mesmos. Deste tratamento resultou um conjunto de SKUs organizados por área e loja, e dentro de cada loja por categoria. Foi também realizado uma agregação temporal dos dados passando os valores registados a serem referentes a vendas semanais por SKU ao invés de vendas diárias por SKU. Esta agregação está associada à frequência temporal com que o retalhista trabalha, ou seja, numa base semanal.

Conforme já foi referido, sendo o objetivo deste trabalho a previsão de procura intermitente, impôs-se obrigatoriamente uma reflexão acerca de quais os métodos adequados para a realização de previsões desse carácter, sendo que existe um espectro considerável de métodos que variam entre métodos muito simples, como o método Naïve, que se baseia na utilização do último valor observado como previsão do valor seguinte, até métodos tão complexos como os modelos de espaço de estados.

Seguidamente são introduzidos os diversos métodos e modelos que foram utilizados no desenvolvimento deste trabalho através duma contextualização desde o seu aparecimento até à atualidade por forma a permitir compreender a evolução de que têm sido alvo, acrescentando ainda exemplos explicativos do funcionamento prático destes métodos. Estes exemplos têm como objetivo colmatar a escassez de demonstração prática do funcionamento destes métodos, pois sendo estes de alguma complexidade a literatura disponível acerca deles carece de demonstração por motivos de simplicidade de exposição dos assuntos em discussão, ficando usualmente apenas disponível ao leitor a explicação teórica que implica uma profunda reflexão posterior acerca do seu funcionamento podendo induzir em erros se a interpretação teórica não for totalmente correta.

¹ Uma série temporal consiste num conjunto de observações de uma variável, feitas em períodos sucessivos de tempo, durante um determinado intervalo.

Começa-se por introduzir o método Naïve que, apesar de ser um dos métodos mais simples e o seu conceito ser do conhecimento geral, este deve ser mencionado uma vez que é recorrentemente referido ao longo deste documento.

2.1.1. Método Naïve

O método Naïve é um método de previsão apropriado para aplicação a séries temporais, que estabelece que o valor previsto para um instante temporal futuro é o último valor observado, conforme mostra a equação seguinte:

$$\hat{y}_t = y_{t-1} \quad (1)$$

Onde y_{t-1} é o valor da série observado no instante $t - 1$ e \hat{y}_t é o valor da previsão para o instante t . Apesar da sua simplicidade não nos deixemos enganar pois este método é ainda muito utilizado nos dias de hoje e inclusive, é um dos mais recorrentes para ser utilizado como referência (*benchmark*). Hyndman et al. (2013) especificam mesmo que este método apresenta uma *performance* bastante interessante em previsões de séries temporais financeiras e económicas.

Existe ainda uma versão do método Naïve específica para dados sazonais em que cada previsão é igual ao último valor observado do respetivo período homólogo. Tanto a aplicação desta versão do método Naïve como a anterior estão demonstradas na figura 6. No método, também ilustrado na figura, a previsão consiste no valor médio de todos os dados históricos da série.

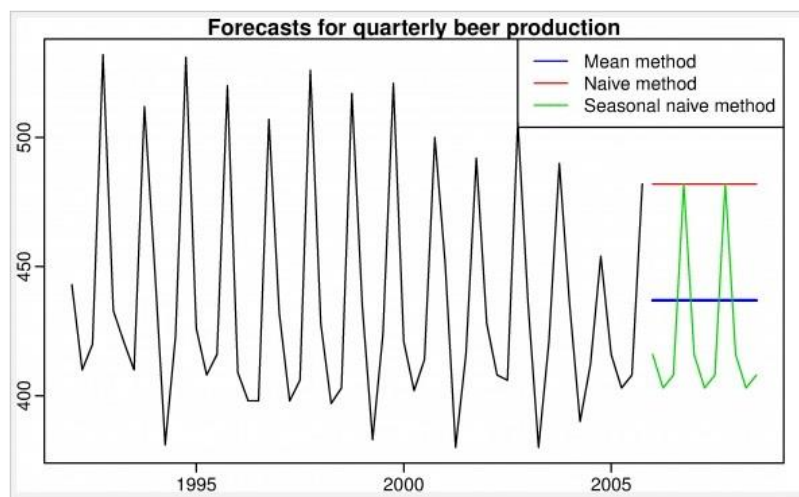


Figura 6 - Previsão utilizando os métodos Naïve, Naïve sazonal e método da média (Hyndman et al, 2013).

Passando à análise dos métodos mais complexos e do seu estado de desenvolvimento até ao momento presente, o Alisamento Exponencial Simples (*Simple Exponential Smoothing* -

SES) foi um dos primeiros a ser aplicado a dados intermitentes surgindo em 1956 pelas mãos de Robert G. Brown (Brown, 1956).

2.1.2. Simple Exponential Smoothing

Fundamentalmente, o SES é um método que determina a previsão para um instante t como a média ponderada entre o valor observado y_{t-1} e o valor previsto \hat{y}_{t-1} mais recente, sendo a sua forma a seguinte:

$$\hat{y}_t = \alpha y_{t-1} + (1 - \alpha) \hat{y}_{t-1} \quad (2)$$

Onde \hat{y}_t é o valor da previsão para o instante de tempo t , y_{t-1} é o valor observado no instante de tempo $t-1$ e \hat{y}_{t-1} o valor da previsão para o instante de tempo $t-1$. O parâmetro α é o parâmetro de alisamento e pode variar entre 0 e 1. A função deste parâmetro é a de atribuir um peso às parcelas que estão envolvidas neste cálculo. Quanto mais próximo de 0 estiver o valor de α (alfa) maior será o peso atribuído às observações mais antigas. Quanto mais próximo de 1 estiver o valor de α maior será o peso atribuído às observações mais recentes. Se apenas a componente mais recente tiver influência na determinação da previsão, o que acontece quando α (alfa) = 1, então a formula transforma-se em $\hat{y}_t = y_{t-1}$ que corresponde ao método *Naïve*. Wallstrom e Segerted (2010) explicitam-no quando demonstram no seu trabalho que quanto maior for o valor que α assume maior é a resposta das previsões às alterações nos dados, contudo menos robustas ao ruído nos mesmos.

O valor a utilizar para α (alfa) pode ser resultado, na forma mais simplificada, de um valor arbitrariamente escolhido através do recurso à experiência, ou pode ser obtido pela utilização de métodos de otimização específicos que permitem obter um valor adequado para este parâmetro de alisamento em função do conjunto de dados em análise. Alguns trabalhos, como Croston (1972), Syntetos et al. (2005) e Teunter e Sani (2009) aconselham que este parâmetro de alisamento tome valores dentro dos limites 0.05 e 0.3, tendo contudo sido demonstrado por Petropoulos et al. (2013) que, nesta temática “parâmetros arbitrados vs parâmetros otimizados”, a seleção dos valores por recurso a processos de otimização permite obter resultados mais exatos e diminui a tendenciosidade dos métodos nesses casos, quando comparados com os mesmos métodos com valores arbitrados. Ou seja, sempre que possível deve-se optar por parâmetros otimizados pois estes demonstram ser menos tendenciosos na previsão que realizam permitindo previsões mais corretas. Kourentzes et al. (2014) demonstram que a *performance* dos métodos quando se recorre à utilização de parâmetros e valores iniciais otimizados resultam em melhorias consideráveis na aplicação do CRO e métodos derivados, tendo ainda sido determinado que nestas condições a *performance* entre os diversos métodos era bastante semelhante entre si mas quando comparada com métodos de parâmetros e valores iniciais não otimizados os primeiros tinham um desempenho superior.

A seleção de parâmetros para a aplicação do SES envolve a inicialização do processo pela determinação de um valor inicial a considerar para o período zero e a otimização do parâmetro de alisamento α (alfa). O valor inicial para o período zero, ou nível l_0 , pode ser determinado de uma de duas formas: utilizar o valor de l_0 igual ao valor real de procura no período 1 ($l_0 = y_1$) ou então utilizar métodos de otimização minimizando a Soma dos Quadrados dos Erros (*Sum of Squared Errors* - SSE):

$$\min SSE = \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \sum_{t=1}^T e_t^2 \quad (3)$$

Este processo baseia-se em recorrer a uma função custo, ou perda, que permite avaliar a adaptação do modelo ao conjunto de dados reais e que, por minimização dessa função se obtenham os valores dos parâmetros do modelo, ou seja, uma otimização dos valores dos parâmetros.

Quanto ao valor de α (alfa), este pode ser estipulado pelo analista ou determinado recorrendo ao processo de otimização por minimização da Soma dos Quadrados dos Erros referido.

Esta métrica do erro é das mais amplamente utilizadas para a otimização de parâmetros em métodos de previsão, porém, é de referir que existem outras funções custo possíveis de serem utilizadas para o efeito sendo que o erro absoluto médio (*Mean Absolute Error* - MAE) começa a ganhar algum destaque como possível substituto do SSE, uma vez que o cálculo do SSE é baseado em erros quadráticos tornando-se sensível a valores extremos, enquanto que o MAE fundamenta-se no cálculo do erro absoluto, eliminando essa influência de extremos:

$$\min MAE = n^{-1} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (4)$$

Correspondendo n ao número de elementos constantes da série temporal de dados.

Atentando ao exemplo presente na Tabela 1, podemos observar o caso de um conjunto de dados de vendas gerados aleatoriamente por recurso ao *software R*, cujos valores de procura contemplam quinze registos que podem ser visualizados na segunda coluna, y_t . Na terceira coluna, que está identificada por SES, é apresentado o valor do nível, l_t . Na quarta coluna temos a previsão, \hat{y}_t , e na última coluna o erro ou resíduo:

$$Resíduo = y_t - \hat{y}_t \quad (5)$$

Por forma a melhor compreender o exemplo da tabela introduz-se ainda uma outra representação possível do SES que é a forma de componente. Para além da forma de média ponderada inicialmente apresentada, uma outra forma possível para o SES é a forma componente que reforça a única componente que existe neste método é o nível da série. A forma de componente deste método é constituída por duas equações que são as seguintes:

$$\text{Equação da previsão} \quad \hat{y}_{t+1|t} = l_t \quad (6)$$

$$\text{Equação do nível} \quad l_t = \alpha y_t + (1 - \alpha) l_{t-1} \quad (7)$$

Na Tabela 1, a terceira coluna, identificada por SES, corresponde à equação do nível (Equação 7), enquanto que a quarta coluna corresponde à equação de previsão (Equação 6) que não é mais do que a simples atribuição do valor da coluna SES do período anterior.

Pensando na seleção dos parâmetros de inicialização, no caso do SES isto implica determinar apenas dois parâmetros. Uma vez que o objetivo deste exemplo é a simples demonstração, explicação e compreensão do funcionamento do SES, os parâmetros iniciais do método foram selecionados da forma simples, ou seja, $l_0 = y_1 = 1.364499190$, conforme explicado anteriormente, e optou-se por α (alfa) = 0.1.

Começando por analisar o ponto a) assinalado na tabela e sombreado a azul, temos que $l_2 = 1.22804927$, sendo que este é o valor de nível no instante 2 e é obtido da forma seguinte

$$l_2 = \alpha y_2 + (1 - \alpha)l_1 = 0.1 \times 0 + (1 - 0.1) \times 1.364499190 = 0.9 \times 1.364499190 = 1.228049271 \quad (8)$$

Passando ao cálculo realizado no ponto b), onde é determinado o valor do erro da previsão para o período quatro, temos que o erro = 1.859350149. Seguindo a fórmula para determinação do erro já introduzida tem-se:

$$Residuo = y_4 - \hat{y}_4 = 2.964594493 - 1.105244344 = 1.859350149 \quad (9)$$

Tabela 1 - Exemplo de aplicação do Simple Exponential Smoothing.

		$\alpha = 0.1$				
Período	y_t	SES = l_t	\hat{y}_t	RESÍDUOS		
0		1,364499190			Ini	
1	1,364499190	1,364499190	NA	NA	$\alpha =$	0,1
2	0	a) 1,228049271	1,364499190	-1,36449919	$l_0 =$	1,36449919
3	0	1,105244344	1,228049271	-1,228049271		
4	2,964594493	1,291179359	1,105244344	b) 1,859350149		
5	0	1,162061423	1,291179359	-1,291179359		
6	0,663400217	1,112195302	1,162061423	-0,498661206		
7	7,202136298	1,721189402	1,112195302	6,089940995		
8	3,533886847	1,902459146	1,721189402	1,812697446		
9	2,29893511	1,942106743	1,902459146	0,396475963		
10	1,826326669	1,930528735	1,942106743	-0,115780074		
11	0	1,737475862	1,930528735	-1,930528735		
12	0	1,563728276	1,737475862	-1,737475862		
13	0,356198355	1,442975284	1,563728276	-1,20752992		
14	0	1,298677755	1,442975284	-1,442975284		
15	2,248060461	1,393616026	1,298677755	0,949382706		
16			c) 1,393616026			
17			1,393616026			
18			1,393616026			
19			1,393616026			
20			1,393616026			

Por fim, temos o exemplo do ponto c) assinalado a azul na Tabela 1 e que está marcado para que seja referenciado que todas as previsões que são feitas correspondem sempre ao último valor de nível que foi calculado, ou seja, no caso do exemplo da tabela o último valor de nível para o período 15, uma vez que o nosso conjunto de dados só abrange o registo de quinze valores, daí que todas as previsões realizadas entre o período dezasseis e o período vinte são iguais entre si e iguais ao valor do último nível determinado, ou seja, 1.393616026.

Este método que acabamos de analisar foi provado ser tendencioso, uma vez que, após registar-se um valor de procura diferente de zero, este método apresenta uma tendência crescente, ou seja, conforme Croston (1972) defendeu no seu trabalho este método possui

uma tendência de ponto de decisão, que é prejudicial para o cálculo de *stocks* que são efetuados neste ponto no cálculo (Nikolopoulos et al. 2011).

Contudo, sendo este um método de previsão considerado simples ainda é possível ver a sua utilização na grande maioria dos trabalhos de investigação e a sua aplicação em muitos *softwares* de previsão utilizados na indústria, precisamente por ser um método de fácil compreensão e aplicação direta aos dados. Apesar da sua simplicidade, e de ter sido já ultrapassado por mais de seis décadas de estudos nesta área, ainda surgem relatos de trabalhos onde este método apresenta uma boa performance. Wallstrom et al. (2010) num trabalho realizado em 2010, bem como Kourentzes (2014) referem que o SES apresenta um comportamento interessante quando comparado com métodos como o CRO e o SBA (explicados de seguida), que são métodos posteriores a este e considerados de maior complexidade.

Na sequência do trabalho de R.G.Brown, surgiu em 1972 o trabalho de J.D.Croston (Croston, 1972) que demonstrou a tendenciosidade do SES e propôs um novo método, que possui o seu nome, e foi posteriormente corrigido por Rao (1973).

2.1.3. Croston

O método de Croston (de seguida designado CRO) surge na tentativa de suplantar a falta de exatidão do SES, e é o método com maior utilização em termos globais na previsão de produtos de procura intermitente.

Assim, a ideia que Croston (1972) teve ao abordar o SES foi a de realizar uma decomposição da série original em duas componentes:

- (1) Os valores de procura não nula (também designados tamanhos da procura);
- (2) Os intervalos de tempo entre procura não nula consecutiva.

A estimação de cada uma destas componentes é realizada separadamente utilizando SES. Assim, a estimativa da componente constituída pelos valores de procura não nula, para o período t é dada por:

$$\hat{z}_t = \hat{z}_{t-1} + \alpha_z(z_{t-1} - \hat{z}_{t-1}), \quad (10)$$

E, a estimativa da componente constituída pelos intervalos de tempo entre procura não nula consecutiva, para o período t é dada por:

$$\hat{p}_t = \hat{p}_{t-1} + \alpha_p(p_{t-1} - \hat{p}_{t-1}), \quad (11)$$

Sendo \hat{z}_t a estimação do valor de procura não nula no período t e z_t o valor de procura não nula efetivamente registado no período t ; \hat{p}_t é a estimação do intervalo de tempo entre procura não nula para o período t e p_t o valor do intervalo de tempo entre procura não nula efetivamente registado no período t .

A figura 7 permite ter uma noção mais concreta da forma como esta decomposição se realiza.

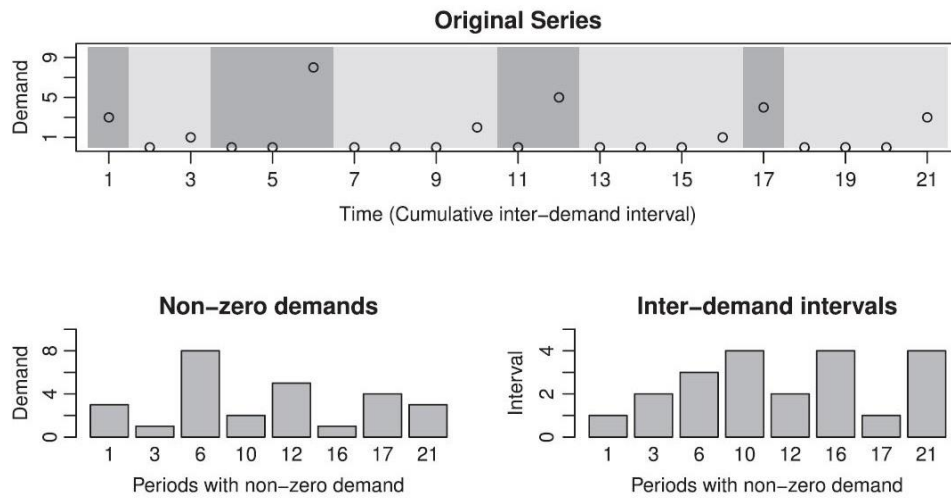


Figura 7 - Exemplo da decomposição de Croston (Petroopoulos, 2016).

A previsão \hat{y}_t é obtida pelo quociente das estimativas destas duas componentes correspondendo ao valor da taxa de procura, sendo para o instante t dada por:

$$\hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t} \quad (12)$$

As estimativas de \hat{z}_t e \hat{p}_t só sofrem atualização quando se regista a ocorrência de procura na série temporal mantendo-se constantes caso contrário, ou seja, quando a procura é zero.

Uma outra forma de apresentar o método de Croston e que torna mais evidente o que acabamos de referir é a presente nas equações (13) e (14):

$$\text{Se } y_t \neq 0 \text{ então } \begin{cases} \hat{z}_t = \hat{z}_{t-1} + \alpha_z(z_{t-1} - \hat{z}_{t-1}) \\ \hat{p}_t = \hat{p}_{t-1} + \alpha_p(p_{t-1} - \hat{p}_{t-1}) \\ \hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t} \end{cases} \quad (13)$$

$$\text{Se } y_t = 0 \text{ então } \begin{cases} \hat{z}_t = \hat{z}_{t-1} \\ \hat{p}_t = \hat{p}_{t-1} \\ \hat{y}_t = \hat{y}_{t-1} \end{cases} \quad (14)$$

Neste modelo estatístico assume-se que a procura ocorre segundo um processo de Bernoulli logo os intervalos entre procura não nula consecutiva têm distribuição geométrica. Assume-se também que os tamanhos de procura não nula são independentes e têm uma distribuição normal.

A utilização do método de Croston envolve a escolha de quatro parâmetros diferentes, nomeadamente dois parâmetros de alisamento distintos, uma vez que existe uma decomposição em duas componentes, consoante se esteja a estimar os valores de procura não nula, recorrendo a α_z , ou se esteja a estimar o tempo entre os valores de procura não nula consecutivos, recorrendo a α_p . Assim, a inicialização do método passa por definir valores para α_z , $l0_z$, α_p e $l0_p$.

A diferenciação dos parâmetros de alisamento, para \hat{z}_t (α_z) e para \hat{p}_t (α_p), permite um aumento da flexibilidade da rapidez de atualização das previsões de cada uma das componentes. Porém, conduz evidentemente a uma complexidade adicional, pelo que recorrentemente, por motivos de simplificação, se utiliza o mesmo valor para ambos os parâmetros de alisamento, $\alpha_z = \alpha_p = \alpha$.

A escolha de valores para estes parâmetros é em tudo similar à exposta para o caso dos parâmetros do SES, sendo que os procedimentos de otimização também se mantêm iguais aos já abordados, sendo então aplicável também neste caso, por exemplo, uma otimização por minimização da Soma dos Quadrados dos Erros.

Para além das funções custo já apresentadas, existe uma outra que tem interesse referir que é o PIS - *Periods In Stock*. O PIS foi proposto por Wallstrom e Segerstedt (2010) e é uma métrica que permite aferir o número de períodos que um SKU esteve em *stock* (ou em *stockout*). Recorre a um *stock* fictício a partir do qual é medida a falta ou o excesso de *stock*, sendo dado por:

$$PIS = - \sum_{i=1}^n \sum_{t=1}^i (y_t - \hat{y}_t) \quad (15)$$

Sendo n o número total de períodos existentes na série temporal e a variável t aquela que nos permite percorrer o somatório entre o primeiro período da série e o período que está atualmente a ser analisado.

A minimização do seu valor absoluto é utilizada para a otimização dos parâmetros iniciais de métodos de previsão, sendo esperado obter no processo de minimização o valor mínimo da tendenciosidade do ajuste do modelo em vez do mínimo do ajuste dos erros.

Esta métrica é também frequentemente utilizada para avaliar a eficiência de um método de previsão para dados intermitentes. O seu interesse está na robustez superior que apresenta quando comparada com as métricas de erro convencionais, devido ao duplo somatório consecutivo que realiza.

Seguidamente, é introduzido um exemplo da aplicação do método de Croston por forma a tornar compreensível a forma como este funciona. Com o objetivo de seguir os passos do método de uma forma mais clara e tornar visualmente menos carregada a tabela e mais facilmente interpretável, fracionou-se a representação do método em três tabelas correspondendo a primeira à determinação da componente \hat{z}_t , a segunda à determinação da componente \hat{p}_t e a terceira à determinação da previsão, por recurso aos valores obtidos nas duas primeiras tabelas, e do resíduo.

A inicialização do CRO presente no exemplo foi realizada sem otimização dos parâmetros uma vez que pretende ser apenas ilustrativa do funcionamento deste método. O conjunto de dados é o mesmo que foi utilizado no exemplo da Tabela 1.

Os parâmetros de alisamento $\alpha_z = \alpha_p = 0.1$ foram escolhidos dentro da gama de valores aconselhados pela literatura, ou seja, $0.1 \leq \alpha_z, \alpha_p \leq 0.3$. Para os valores iniciais $l0_z$ e $l0_p$ optou-se pelo mesmo procedimento do exemplo do SES e igualou-se o valor inicial ao primeiro valor registado, ou seja, $l0_z = z_1$ enquanto que $l0_p = p_1$.

Tabela 2- Determinação da componente \hat{z}_t do método de Croston.

		$\alpha = 0.1$				
Período	y_t	z_t	SES	\hat{z}_t		
0		----	1,364499190		Ini	
1	1,36449919	1,364499190	1,364499190	NA	$\alpha =$	0,1
2	0		1,364499190	1,364499190	$l_{0_z} =$	1,364499190
3	0		a) 1,364499190	1,364499190	$l_{0_p} =$	1,000000000
4	2,964594493	2,964594493	b) 1,524508720	1,364499190		
5	0	c)	1,524508720	1,524508720		
6	0,663400217	d) 0,663400217	1,438397870	1,524508720		
7	7,202136298	7,202136298	2,014771713	1,438397870		
8	3,533886847	3,533886847	2,166683226	2,014771713		
9	2,29893511	2,298935110	2,179908414	2,166683226		
10	1,826326669	1,826326669	2,144550240	e) 2,179908414		
11	0		2,144550240	2,144550240		
12	0		2,144550240	2,144550240		
13	0,356198355	0,356198355	1,965715051	2,144550240		
14	0		1,965715051	1,965715051		
15	2,248060461	2,248060461	1,993949592	1,965715051		
16				1,993949592		
17				1,993949592		
18				1,993949592		
19				1,993949592		
20				1,993949592		

Analisando a informação constante na Tabela 2, temos na primeira coluna os períodos de tempo analisados; na segunda coluna a série temporal com os valores de procura; na terceira coluna a componente z_t dos valores de procura não nula; na quarta coluna o nível l_z , que é determinado por recurso ao SES conforme demonstrado atrás; e por fim, na última coluna, a previsão para cada um dos períodos em análise.

Começando por analisar os pontos a) e b), o primeiro ponto encontra-se marcado para reforçar uma característica deste método que é a de só se realizar atualização dos valores de tamanho da procura quando ocorre um valor de procura diferente de zero, daí que, como a procura é zero nos períodos 2 e 3 o valor de a) mantém-se igual aos anteriores. No caso do ponto b), instante em que já temos um valor de procura não nulo, este é estimado pelo SES da forma apresentada na equação seguinte:

$$l_4 = \alpha z_4 + (1 - \alpha)l_3 = 0.1 \times 2.964594493 + (1 - 0.1) \times 1.364499190 = 1.524508720 \quad (16)$$

Comentando agora os pontos c) e d) da Tabela 2, verifica-se que em c) não existe registo de procura o que ocorre pela ausência de procura não nula na série temporal. No ponto d) é demonstrado que os valores que constituem a componente de tamanho de procura correspondem aos valores de procura registados com a exclusão das observações em que se regista procura zero.

O ponto e) corresponde à determinação da previsão da componente em cálculo e assume o valor igual ao nível do período anterior, ou seja, neste caso em específico $\hat{z}_{10} = l_9$, obtendo-se no final de todos os períodos a série completa para a componente \hat{z}_t .

Passando agora à análise da Tabela 3, desta consta a determinação da componente de tempo entre valores consecutivos de procura diferente de zero.

Tabela 3 - Determinação da componente \hat{p}_t do método de Croston.

		$\alpha = 0.1$				
Período	y_t	p_t	SES	\hat{p}_t		
0		----	1		Ini	
1	1,36449919	1	1,000000000	NA	$\alpha =$	0,1
2	0		1,000000000	1,000000000	$l_{0_z} =$	1,364499190
3	0		1,000000000	1,000000000	$l_{0_p} =$	1,000000000
4	2,964594493	a) 3	1,200000000	1,000000000		
5	0	b)	1,200000000	1,200000000		
6	0,663400217	2	1,280000000	1,200000000		
7	7,202136298	1	c) 1,252000000	1,280000000		
8	3,533886847	1	1,226800000	1,252000000		
9	2,29893511	1	1,204120000	d) 1,226800000		
10	1,826326669	1	1,183708000	1,204120000		
11	0		1,183708000	1,183708000		
12	0		1,183708000	1,183708000		
13	0,356198355	3	1,365337200	1,183708000		
14	0		1,365337200	1,365337200		
15	2,248060461	2	1,428803480	1,365337200		
16				1,428803480		
17				1,428803480		
18				1,428803480		
19				1,428803480		
20				1,428803480		

As duas primeiras colunas são iguais à Tabela 2; a terceira coluna refere-se ao tempo que decorre entre procuras consecutivas que sejam diferentes de zero e esta é apenas alvo de atualização nos momentos em que ocorre procura não nula, p_t ; na quarta encontra-se o nível do intervalo entre procuras consecutivas, l_p ; por fim, na quinta coluna é apresentada a previsão associada a esta componente, \hat{p}_t .

Começando por analisar os pontos a) e b) da Tabela 3, no primeiro ponto, que é apresentado num período com procura não nula após dois zeros de procura, temos o espaço que decorre entre períodos de procura diferentes de zero, sendo a distância de 3 unidades temporais uma vez que temos dois períodos com zero (período 2 e 3) e apenas no terceiro período temporal (período 4) é que ocorre um valor de procura diferente de zero. No segundo ponto, b), não se verifica nenhum registo e esse facto deve-se ao conjunto de dados apresentar como registo de procura um zero, pelo que não existe atualização do valor do intervalo de tempo entre procuras consecutivas.

O ponto marcado na Tabela 3 como c) corresponde ao cálculo da equação de nível para o período 7, l_7 , e é determinado da forma seguinte:

$$l_7 = \alpha \times p_7 + (1 - \alpha) \times l_6 = 0.1 \times 1 + (1 - 0.1) \times 1.28 = 1.252 \quad (17)$$

O último ponto, d), apresenta a determinação da previsão da componente do intervalo de tempo que consiste no nível do instante anterior, ou seja, neste caso, $\hat{p}_9 = l_8$.

A última tabela relativa a este exemplo (Tabela 4) apresenta o cálculo da previsão final e a determinação dos resíduos, ou erros, obtidos em cada período.

Por forma a visualizarmos mais concretamente o que acontece nesta fase da aplicação do método são apresentados dois exemplos, o ponto a) e o ponto b).

O ponto a) apresenta a determinação do valor de previsão para o período 5. Este valor é determinado, segundo a fórmula já apresentada, e o seu cálculo é realizado seguindo a equação (18).

$$\hat{y}_5 = \frac{\hat{z}_5}{\hat{p}_5} = \frac{1.524508720}{1.2} = 1.270423933 \quad (18)$$

No ponto b) é calculado o resíduo referente à previsão feita no período 6, sendo este dado no caso do instante 6 por:

$$Erro = y_6 - \hat{y}_6 = 0.663400217 - 1.270423933 = -0.607023716 \quad (19)$$

Conforme podemos verificar pelo exemplo apresentado, a aplicação do método de Croston é relativamente fácil de ser compreendida quando posta em prática.

Este apresenta-se como um método interessante pela simplicidade de compreensão tendo, porém, já sido demonstrado que possui algumas características que o tornam menos adequado para a previsão de procura intermitente.

Tabela 4 - Determinação da previsão e resíduo do método de Croston.

		$\alpha = 0.1$			
Período	y_t	\hat{y}_t	RESÍDUOS	Ini	
0				$\alpha =$	0,1
1	1,36449919	NA	NA	$l0_z =$	1,364499190
2	0	1,36449919	-1,36449919	$l0_p =$	1,000000000
3	0	1,36449919	-1,36449919		
4	2,964594493	1,36449919	1,600095303		
5	0	a) 1,270423933	-1,270423933		
6	0,663400217	1,270423933	b) -0,607023716		
7	7,202136298	1,123748336	6,078387962		
8	3,533886847	1,609242582	1,924644265		
9	2,29893511	1,766125877	0,532809233		
10	1,826326669	1,810374725	0,015951943		
11	0	1,81172235	-1,81172235		
12	0	1,81172235	-1,81172235		
13	0,356198355	1,81172235	-1,455523995		
14	0	1,439728626	-1,439728626		
15	2,248060461	1,439728626	0,808331835		
16		1,395538029			
17		1,395538029			
18		1,395538029			
19		1,395538029			
20		1,395538029			

2.1.4. Aproximação de Syntetos-Boylan

Em 2001 foi provado por Syntetos et al. (2001) que o método de Croston é enviesado, e que o enviesamento está relacionado com a forma como é determinada a previsão.

Estes investigadores concluíram que as estimativas realizadas em separado para o tamanho de procura e para o intervalo entre procuras consecutivas não nulas não apresentam enviesamento. E que o enviesamento do método de Croston tem origem no rácio utilizado para determinar a previsão resultante da combinação destas duas componentes. Simultaneamente estes investigadores tentaram relacionar a variação do enviesamento do método de Croston com a variação dos valores do parâmetro de alisamento α (alfa). Ficou provado que o enviesamento do método aumenta com o aumento dos valores de α (alfa) não tendo estes autores conseguido identificar uma relação específica entre estas quantidades. Segundo estes investigadores o valor do enviesamento é máximo quando α (alfa) = 1 podendo o seu valor ser determinado através da equação seguinte:

$$\mu \left[-\frac{1}{p-1} \log \frac{\mu}{p} \right] - \frac{\mu}{p} \quad (20)$$

Sendo, μ a média dos valores e p é o intervalo médio entre valores de procura consecutivos.

Neste trabalho ficou também provado que, para valores de α (alfa) superiores a 0.15 o enviesamento das previsões tornar-se pronunciado, daí existir a perspetiva generalizada de utilizar valores baixos para os parâmetros de alisamento, quando estes não são otimizados.

Com o objetivo de solucionar o problema do enviesamento do método de Croston, foi apresentada por Syntetos et al. (2005) uma adaptação deste método que ficou conhecida como aproximação de Syntetos e Boylan (Syntetos-Boylan Approximation - SBA).

O SBA é uma adaptação do modelo matemático do método de Croston original que introduz um fator multiplicativo na equação da previsão, permitindo que esta seja aproximadamente não-enviesada. Ou seja, consiste em multiplicar o valor da previsão pelo fator de amortecimento $(1 - \alpha_p/2)$. Segundo estes investigadores a nova previsão é aproximadamente não enviesada uma vez que

$$\begin{aligned} E \left[\left(1 - \frac{\alpha}{2} \right) \left(\frac{Z_t}{p_t} \right) \right] &= \left(1 - \frac{\alpha}{2} \right) E \left[\frac{Z_t}{p_t} \right] \\ &= \left(1 - \frac{\alpha}{2} \right) \left(\frac{\mu}{p} + \frac{1}{2} \frac{\partial^2 \left(\frac{\mu}{p} \right)}{\partial p^2} \text{Var}(p) \right) \\ &= \left(\frac{2-\alpha}{2} \right) \left(\frac{\mu}{p} + \frac{\alpha}{2-\alpha} \mu \frac{p-1}{p^2} \right) \\ &= \frac{\mu}{p} \left(\frac{2-\alpha}{2} + \frac{\alpha}{2} \frac{p-1}{p} \right) \approx \frac{\mu}{p} \end{aligned} \quad (21)$$

Assim, Syntetos et al. (2005) propõem que a previsão seja dada pela seguinte equação:

$$\hat{y}_t = \left(1 - \frac{\alpha_p}{2} \right) \frac{\hat{z}_t}{\hat{p}_t} \quad (22)$$

Apesar de o SBA ser uma modificação interessante do CRO, Teunter et al. (2009b) e Wallstrom et al. (2010) demonstram que este método continua a apresentar enviesamento nas previsões geradas sobretudo quando a série temporal em análise apresenta baixa intermitência, ou seja, quando a quantidade de zeros é reduzida.

Shale et al. (2006) demonstraram que a metodologia proposta por Syntetos et al. (2005) sofreria a modificação que se pode observar na equação (23) quando a procura ocorre segundo um processo de *Poisson*:

$$\hat{y}_t = \left(1 - \frac{\alpha}{2-\alpha} \right) \frac{\hat{z}_t}{\hat{p}_t} \quad (23)$$

Neste trabalho Shale et al. (2006) sugerem ainda substituir o SES, utilizado para prever as componentes da série temporal original, por uma média móvel (MA) de ordem k , resultando a equação da previsão seguinte:

$$\hat{y}_t = \left(\frac{k}{k+1} \right) \frac{\hat{z}_t}{\hat{p}_t} \quad (24)$$

A mesma alteração do SES pelo MA(k) no caso da procura ocorrer segundo um processo de Poisson resulta na equação seguinte:

$$\hat{y}_t = \left(\frac{k-1}{k} \right) \frac{\hat{z}_t}{\hat{p}_t} \quad (25)$$

Kourentzes (2013) apresenta ainda um modelo que segue igualmente a filosofia Croston e que apesar de ser frequentemente utilizado é raramente explorado na literatura. Este modelo realiza a estimação de z_t e y_t com um *random walk* utilizando α (alfa) = 1 ou $k = 1$ nos métodos apresentados. Fundamentalmente, neste caso o método baseia-se num *Naïve*.

Apesar da sua simplicidade, quando comparado com os restantes métodos que aqui são apresentados, o método *Naïve* é considerado um excelente *benchmark* (Babai et al, 2011; Petropoulos et al, 2014), sendo referenciado como um bom método de previsão principalmente no caso de séries temporais com elevado grau de intermitência, uma vez que quando temos procuras nulas consecutivas este método irá conseguir uma previsão exata do valor de consumo, obtendo um erro zero. Contudo, conforme Petropoulos et al. (2015) reforçam, esta característica de boa previsibilidade do *Naïve* em séries altamente intermitentes, não tem lógica na perspetiva empresarial, uma vez que erro zero numa ocorrência de consumo zero implica zero encomendas e manutenção de *stock*.

O processo de inicialização do SBA envolve definir dois parâmetros de alisamento e dois valores iniciais, respetivamente, α_z , α_p , $l0_z$ e $l0_p$. De forma similar aos métodos anteriormente apresentados a seleção destes parâmetros pode ser feita de uma forma simples fazendo-se $\alpha_z = \alpha_p = \alpha$, ou então pode ser realizado um processo de otimização destas variáveis recorrendo-se à minimização da Soma dos Quadrados dos Erros, tendo sido demonstrado empiricamente por Petropoulos et al. (2013) as melhorias que a aplicação desta metodologia de seleção de parâmetros trás em termos de enviesamento das previsões obtidas.

Syntetos et al. (2005) propõem que se utilize para estes parâmetros α_z , α_p o intervalo de valores [0.05, 0.2]. Teunter et al. (2010) e Snyder (2002) realizaram estudos acerca das escolhas ideais para estes parâmetros continuando-se, porém, a utilizar frequentemente escolhas *ad hoc* (Syntetos e Boylan, 2005; Teunter e Duncan, 2009; Romeijnders et al., 2012) alegando-se que o processo de otimização não é simples devido à frequente falta de observações de procura diferentes de zero no conjunto de dados disponíveis, sendo recorrente realizar-se a estimação das componentes da série recorrendo a um único parâmetro de alisamento, ou seja, $\alpha_z = \alpha_p$. Quanto à seleção dos valores iniciais, ou seja, $l0_z$ e $l0_p$ a opção mais simplista continua a passar pelo recurso à utilização do valor observado no primeiro período das componentes resultantes da decomposição da série original (com procura não nula), ou seja, $l0_z = z_1$ e $l0_p = p_1$. A otimização dos valores iniciais passa também pela minimização do SSE.

Antes de passarmos à análise de um exemplo demonstrativo de aplicação do SBA, introduz-se uma outra função custo recorrentemente utilizada na otimização dos parâmetros e valores iniciais destes métodos. Essa função foi proposta por Kourentzes (2014) e é conhecida por MAR (*Mean Absolute Rate*), tendo sido demonstrado por este que conduz a previsões mais corretas quando comparada com as outras funções custo já referidas:

$$\min MAR = \sum_{i=1}^n |\hat{y}_i - i^{-1} \sum_{j=1}^i y_j| \quad (26)$$

Nesta secção apresenta-se ainda um exemplo da aplicação do SBA, presente nas Tabelas 5 e 6, que permitirá obter uma perspetiva mais concreta do seu funcionamento. A inicialização realizada para efeitos de demonstração foi a mais simplificada, ou seja, $\alpha_z = \alpha_p = 0.1$ e $l0_z = z_1 = 1.364499190$ e $l0_p = p_1 = 1$.

As Tabelas 5 e 6 estão estruturadas de forma análoga às tabelas (2, 3 e 4) apresentadas no exemplo de aplicação do método Croston. A alteração que se verifica é na formula de cálculo da previsão, \hat{y}_t , sendo que de resto podemos encontrar na primeira tabela a determinação da quantidade de procura (\hat{z}_t) por recurso à determinação da equação de nível e recurso ao SES e na segunda tabela temos a determinação do intervalo entre procuras diferentes de zero (\hat{p}_t) calculado também por recurso ao SES e, por fim, nas duas últimas colunas encontramos a determinação da previsão de procura, resultante do rácio das duas componentes já calculadas afetadas do fator proposto por Syntetos, e o cálculo do erro associado às previsões em cada período. Uma vez que a exemplificação da determinação das duas componentes já foi demonstrada para o método de Croston, e o SBA é realizado exatamente da mesma forma, no presente exemplo remetemos para as demonstrações realizadas para os pontos de a) a e) da Tabela 2 e as demonstrações realizadas para os pontos de a) a d) da Tabela 3.

Em relação ao ponto a) da Tabela 6, nesta coluna são apresentadas as previsões realizadas pelo SBA. O valor obtido nesse ponto recorre ao resultado obtido na linha 8 da coluna \hat{z}_t da Tabela 5 e à linha 8 da coluna \hat{p}_t da Tabela 6 e realiza o cálculo do rácio entre estas duas componentes multiplicado pelo novo fator introduzido, ou seja:

$$\hat{y}_8 = \left(1 - \frac{\alpha}{2}\right) \times \frac{\hat{z}_8}{\hat{p}_8} = \left(1 - \frac{0.1}{2}\right) \times \frac{2.014771713}{1.2520} = 1.528780453 \quad (27)$$

Se refletirmos por breves momentos no valor previsto com o SBA para o período 8 e a correspondente previsão realizada com o Croston, uma vez que a série temporal utilizada em ambos os exemplos é a mesma, e comparando com a procura real verificada, conseguimos obter uma ideia do comportamento de um método relativamente ao outro para a previsão daquele ponto em específico. Para esta série temporal específica verifica-se que o Croston apresenta um resíduo inferior para o mesmo período quando comparado com o SBA.

Tabela 5 - Determinação da componente \hat{z}_t do SBA.

Período	y_t	z_t	SES	\hat{z}_t		
0		----	1,364499190		Ini	
1	1,36449919	1,364499190	1,364499190	NA	$\alpha =$	0,1
2	0		1,364499190	1,364499190	$l0_z =$	1,364499190
3	0		1,364499190	1,364499190	$l0_p =$	1,000000000
4	2,964594493	2,964594493	1,524508720	1,364499190		
5	0		1,524508720	1,524508720		
6	0,663400217	0,663400217	1,438397870	1,524508720		
7	7,202136298	7,202136298	2,014771713	1,438397870		
8	3,533886847	3,533886847	2,166683226	2,014771713		
9	2,29893511	2,298935110	2,179908414	2,166683226		
10	1,826326669	1,826326669	2,144550240	2,179908414		
11	0		2,144550240	2,144550240		
12	0		2,144550240	2,144550240		
13	0,356198355	0,356198355	1,965715051	2,144550240		
14	0		1,965715051	1,965715051		
15	2,248060461	2,248060461	1,993949592	1,965715051		
16				1,993949592		
17				1,993949592		
18				1,993949592		
19				1,993949592		
20				1,993949592		

Passando à análise do ponto b), neste ponto é determinado o erro associado à previsão realizada no período 12. A fórmula de cálculo do erro é igual à apresentada no método de Croston, baseando-se na diferença entre o valor de procura real e o valor previsto pelo método. O ponto c) demonstra que para períodos posteriores ao último valor de procura real do conjunto de dados os valores previstos pelo método são todos iguais entre si, ou seja, a previsão é sempre a mesma.

Conforme é possível verificar pela exemplificação do funcionamento do método este é de simples aplicação tornando-o assim um método com potencial para ser amplamente empregue em meio empresarial. Ainda mais é reforçado em múltiplos estudos (Eaves et al., 2004; Syntetos et al. 2005, 2006; Teunter et al., 2009; Petropoulos et al., 2013; Petropoulos et al., 2015) que de entre as diversas modificações propostas ao método de Croston este método é aquele que apresenta superior suporte empírico e superior desempenho em previsões quando comparado com o método original.

Tabela 6 - Determinação da componente \hat{p}_t , previsão e erro do SBA.

t	y_t	p_t	SES	\hat{p}_t	\hat{y}_t	RESÍDUOS
0		----	1			
1	1,36449919	1	1,000000000	NA	NA	NA
2	0		1,000000000	1,000000000	1,29627423	-1,29627423
3	0		1,000000000	1,000000000	1,29627423	-1,29627423
4	2,964594493	3	1,200000000	1,000000000	1,29627423	1,668320263
5	0		1,200000000	1,200000000	1,206902737	-1,206902737
6	0,663400217	2	1,280000000	1,200000000	1,206902737	-0,54350252
7	7,202136298	1	1,252000000	1,280000000	1,067560919	6,134575379
8	3,533886847	1	1,226800000	1,252000000	a) 1,528780453	2,005106394
9	2,29893511	1	1,204120000	1,226800000	1,677819583	0,621115526
10	1,826326669	1	1,183708000	1,204120000	1,719855989	0,10647068
11	0		1,183708000	1,183708000	1,721136233	-1,721136233
12	0		1,183708000	1,183708000	1,721136233	b) -1,721136233
13	0,356198355	3	1,365337200	1,183708000	1,721136233	-1,364937877
14	0		1,365337200	1,365337200	1,367742195	-1,367742195
15	2,248060461	2	1,428803480	1,365337200	1,367742195	0,880318266
16				1,428803480	1,325761128	
17				1,428803480	c) 1,325761128	
18				1,428803480	1,325761128	
19				1,428803480	1,325761128	
20				1,428803480	1,325761128	

Na tabela 6 a variável t é representada por motivo de simplificação da tabela e representa os períodos.

Na sequência do seu trabalho com séries temporais intermitentes, aprofundando o método de Croston e posteriores variações, Teunter et al. (2011) apresentam uma variação deste cujo nome que atribuíram foi as iniciais dos três autores, ou seja, o método passou a ser conhecido por TSB.

2.1.5. Teunter, Syntetos e Babai

Existindo já à data alguns investigadores a desenvolver estudos que se centravam na temática do enviesamento do método de Croston, alguns dos quais exploravam novas adaptações por forma a corrigir esse enviesamento, Teunter et al. (2011) decide analisar o CRO numa ótica que ele via ser necessária aprofundar e desenvolver e que era diferente daquela que era o foco da maioria dos investigadores, surgindo então com a perspetiva da necessidade de introduzir informação acerca da obsolescência de produtos nos métodos de previsão, algo que até ao momento não estava explorado nas diversas variantes que surgiam, tão pouco no próprio CRO.

Esta abordagem baseia-se num interesse real e prático do retalhista que é o de ter um método que permite detetar quando um item deixa de ter qualquer procura permitindo-lhe, no menor espaço de tempo possível, tomar decisões, por exemplo, acerca da manutenção ou extinção da venda desse SKU.

Esta introdução do conhecimento acerca da obsolescência dos produtos no método é conseguida através da realização de atualizações num dos componentes em cada período, contrariamente ao que é feito nos métodos propostos anteriormente em que se realiza uma atualização dos componentes apenas nas ocorrências de procura diferente de zero.

Teunter et al. (2011) mantêm a similaridade ao CRO decompondo a série original em duas componentes. A primeira componente, como habitualmente, corresponde ao tamanho da procura z_t e é estimada da forma habitual usando SES. Porém, a segunda componente, que corresponde à probabilidade de procura d_t , é 1 ou 0 consoante ocorre ou não procura no instante t , respetivamente. Esta componente é também estimada recorrendo ao SES e atualizada no final de cada período.

Teunter et al. (2011) introduzem uma separação clara entre o parâmetro de alisamento utilizado no SES da componente z_t , e o parâmetro de alisamento utilizado no SES da componente d_t , neste caso β , existindo assim uma clara necessidade de taxas de atualização diferentes para estas duas componentes. Neste estudo Teunter et al. (2011) concluem a necessidade de que na seleção dos parâmetros de suavização $\alpha_z > \beta$, sendo sempre reforçada a importância da cuidada seleção destes parâmetros para a boa performance do método.

Ao introduzir uma atualização da probabilidade de procura em cada período temporal, este método permite que o valor da previsão se aproxime de zero quando se verificam períodos continuados de procura nula, sendo este facto uma vantagem face às restantes versões do CRO, pois permite olhar para os valores obtidos e tomar decisões mais concretas acerca da gestão de *stocks*.

O TSB é um método que apresenta quatro valores que necessitam de inicialização para que se possa aplicar o método. A inicialização deste método envolve selecionar um nível inicial e um parâmetro de alisamento para a componente do tamanho de procura; e envolve selecionar um nível inicial e um parâmetro de alisamento para a componente da probabilidade de procura.

No que concerne à seleção dos valores iniciais, e de forma similar ao que acontece com os restantes métodos, é possível optar por uma otimização destes parâmetros através da utilização de uma função custo ou então pode realizar-se uma inicialização simples que se baseia em igualar os valores iniciais ao primeiro valor da componente em análise. Em relação à inicialização dos parâmetros de alisamento, o processo de otimização mantém-se possível sendo, contudo, necessário ter em atenção que $\alpha_z > \beta$, preocupação que também deve ser mantida quando a seleção dos parâmetros é feita pelo próprio analista, que é a opção mais utilizada devido à sua simplicidade. O intervalo de valores para estas variáveis mais amplamente defendido é [0.05, 0.3] (Teunter et al., 2011).

A determinação da obsolescência dos SKUs em análise pelo TSB sustenta-se no estabelecimento de níveis limite para os valores previstos para a procura, que sendo ultrapassados uma vez ou continuamente, em função do que é definido, considera-se que esse item passa a ser considerado obsoleto.

A previsão da procura deste método, \hat{y}_t , determina-se através do produto da quantidade de procura estimada pela probabilidade de procura estimada, em cada período. As equações que traduzem o método TSB são as seguintes:

$$\text{Se } d_t = 1 \text{ então } \begin{cases} \hat{d}_t = \beta d_{t-1} + (1 - \beta) \hat{d}_{t-1} \\ \hat{z}_t = \hat{z}_{t-1} + \alpha_z (z_{t-1} - \hat{z}_{t-1}) \\ \hat{y}_t = \hat{d}_t \hat{z}_t \end{cases} \quad (28)$$

$$\text{Se } d_t = 0 \text{ então } \begin{cases} \hat{d}_t = \beta d_{t-1} + (1 - \beta) \hat{d}_{t-1} \\ \hat{z}_t = \hat{z}_{t-1} \\ \hat{y}_t = \hat{d}_t \hat{z}_t \end{cases} \quad (29)$$

Para compreender de forma mais sistemática o funcionamento deste método introduz-se um exemplo ilustrativo que é apresentado nas Tabelas 7 e 8.

Para efeitos de exemplificação utilizaram-se valores iniciais selecionados de forma *ad hoc* sendo $l_0 = y_1$, $\alpha_z = 0.15$, $l_0 = d_1$ e $\beta = 0.1$.

Começando por realizar uma análise geral da informação apresentada na Tabela 7, desta consta, na primeira coluna a informação do período; na segunda o conjunto de dados relativos à procura registada; na terceira o valor da procura registada nos períodos em que a procura é diferente de zero, ignorando-se os períodos em que a procura é igual a zero; na quarta é aplicada a equação de nível de cada período; na quinta é determinada a série de previsão da quantidade de procura.

Em termos de metodologia para obter os valores que estão registados nesta tabela, uma vez que esta apresenta a mesma estrutura que o método de Croston, é redundante exemplificar como se obtiveram os valores aí presentes remetendo tais explicações para a Tabela 2, sendo porém necessário ter em consideração que o valor de α_z nesta tabela é diferente do utilizado no exemplo anterior.

No que concerne à Tabela 8, esta apresenta primeiramente os períodos e o conjunto de dados de que dispomos; de seguida apresenta a probabilidade de procura; na coluna seguinte é utilizada a equação de nível para cada período em questão por recurso ao SES; seguidamente, é aplicada a equação de previsão, ou seja, é determinada a previsão da probabilidade de procura para cada período; por fim é apresentada a previsão da procura e é calculado o erro cometido nessa previsão.

Analisando a aplicação do método para obter a Tabela 8, começando pelo a), neste ponto é apresentada a probabilidade de procura igual a 1, uma vez que nesse instante de tempo ocorreu procura. No caso de a procura ser zero, o valor da probabilidade de procura seria também zero.

No ponto b) determinamos o nível do período 3 por recurso ao SES, em função do nível do período anterior e da probabilidade de procura do próprio nível. Assim, por aplicação do SES às variáveis obtemos a equação (30):

$$l_3 = \beta d_3 + (1 - \beta)l_2 = 0.1 \times 0 + (1 - 0.1)0.9 = 0.81 \quad (30)$$

Tabela 7 - Determinação da componente \hat{z}_t do TSB.

Periodo	y_t	z_t	SES	\hat{z}_t		
0		----	1,364499190		Ini	
1	1,36449919	1,364499190	1,364499190	NA	$\alpha_z =$	0,15
2	0		1,364499190	1,364499190	$l0_z =$	1,364499190
3	0		1,364499190	1,364499190		
4	2,964594493	2,964594493	1,604513485	1,364499190	Ini	
5	0		1,604513485	1,604513485	$\beta =$	0,10
6	0,663400217	0,663400217	1,463346495	1,604513485	$l0_d =$	1
7	7,202136298	7,202136298	2,324164965	1,463346495		
8	3,533886847	3,533886847	2,505623248	2,324164965		
9	2,29893511	2,298935110	2,474620027	2,505623248		
10	1,826326669	1,826326669	2,377376023	2,474620027		
11	0		2,377376023	2,377376023		
12	0		2,377376023	2,377376023		
13	0,356198355	0,356198355	2,074199373	2,377376023		
14	0		2,074199373	2,074199373		
15	2,248060461	2,248060461	2,100278536	2,074199373		
16				2,100278536		
17				2,100278536		
18				2,100278536		
19				2,100278536		
20				2,100278536		

Conforme podemos verificar pela aplicação do SES, apesar de existir um ponto no conjunto de dados de procura igual a zero, este método realiza uma atualização período a período da estimação da probabilidade de procura, o que não aconteceria se este cálculo fosse realizado para o mesmo ponto com um outro método como, por exemplo, o método de Croston, onde se iria manter a previsão anterior numa situação de procura igual a zero.

O ponto c) é referente à obtenção da previsão de probabilidade de procura para o período 7, conforme demonstra a equação (31):

$$\hat{d}_7 = l_6 = 0.77149 \quad (31)$$

Por fim, nos pontos d) e e) são apresentados, respetivamente, o cálculo da previsão de procura e o cálculo do erro de previsão para os respetivos períodos.

Tabela 8 - Exemplo de aplicação do TSB, Determinação de d_t , previsão e erro

t	y_t	d_t	SES	\hat{d}_t	\hat{y}_t	RESÍDUOS
0		----	1			
1	1,36449919	a) 1	1,000000000	NA	NA	NA
2	0	0	0,900000000	1,000000000	1,36449919	-1,36449919
3	0	0	b) 0,810000000	0,900000000	1,228049271	-1,228049271
4	2,964594493	1	0,829000000	0,810000000	1,105244344	1,859350149
5	0	0	0,746100000	0,829000000	1,330141679	-1,330141679
6	0,663400217	1	0,771490000	0,746100000	1,197127511	-0,533727294
7	7,202136298	1	0,794341000	c) 0,771490000	1,128957187	6,07317911
8	3,533886847	1	0,814906900	0,794341000	1,846179523	1,687707324
9	2,29893511	1	0,833416210	0,814906900	d) 2,041849673	0,257085436
10	1,826326669	1	0,850074589	0,833416210	2,062388444	-0,236061775
11	0	0	0,765067130	0,850074589	2,020946946	-2,020946946
12	0	0	0,688560417	0,765067130	1,818852251	e) -1,818852251
13	0,356198355	1	0,719704375	0,688560417	1,636967026	-1,280768671
14	0	0	0,647733938	0,719704375	1,492810364	-1,492810364
15	2,248060461	1	0,682960544	0,647733938	1,343529328	0,904531134
16				0,682960544	1,434407372	
17				0,682960544	1,434407372	
18				0,682960544	1,434407372	
19				0,682960544	1,434407372	
20				0,682960544	1,434407372	

Correspondendo, nesta tabela, t ao período e sendo utilizado para efeitos de simplificação da apresentação da tabela em questão. A previsão de procura, d), é determinada pela multiplicação das estimativas das duas componentes. Este é outro ponto que diferencia o TSB dos seus antecessores. A equação (32) demonstra o cálculo dessa previsão para o período 9:

$$\hat{y}_9 = \hat{z}_9 \times \hat{d}_9 = 2.505623248 \times 0.8149069 = 2.041849673 \quad (32)$$

Analisando a coluna de previsão de procura conseguimos reconhecer a atualização de previsão em cada período, mesmo quando a procura é igual a zero, o que caracteriza este método.

No último ponto, e), é realizada a avaliação da performance da previsão realizada pelo TSB para o período 12 do conjunto de dados. A determinação deste erro é feita de igual forma aos que já foram apresentados em exemplos anteriores:

$$Erro = y_{12} - \hat{y}_{12} = 0 - 1.818852251 = -1.818852251 \quad (33)$$

A obsolescência dos produtos no estudo da previsão de procura é um tema que continuou a ser alvo de estudo e aprofundamento, surgindo recentemente novas propostas de métodos que se debruçam sobre esta capacidade de detetar quando um produto deixa de ser alvo de procura. A determinação da obsolescência da qual um produto pode ser alvo, no caso do TSB, é detetável por recurso à variável d_t , que corresponde à probabilidade de ocorrência de procura. É definido um nível de *threshold* que quando é ultrapassado por esta variável é considerado que o SKU em análise passou a ser caracterizado como obsoleto. Esta informação é fornecida ao retalhista que deverá adotar as decisões necessárias em conformidade com os dados que lhe são fornecidos acerca desse SKU.

2.1.6. Intermittent Multiple Aggregation Prediction Algorithm

A agregação temporal é uma das estratégias utilizadas com objetivo de melhorar a performance dos métodos de previsão. Esta estratégia baseia-se na agregação temporal dos dados disponíveis com o objetivo de, no caso da intermitência, reduzir ou idealmente eliminar a incidência de zeros de procura nos dados disponíveis por forma a viabilizar uma melhoria nos resultados das análises realizadas. Esta estratégia é recorrentemente utilizada por analistas.

O *Intermittent Multiple Aggregation Prediction Algorithm* (iMAPA) é a versão do *Multiple Aggregation Prediction Algorithm* (MAPA) para dados intermitentes. Fundamentalmente, o MAPA obtém as previsões combinando as componentes de estado (nível, tendência e sazonalidade) das séries temporais determinadas com diferentes níveis de agregação temporal. Este método utiliza o facto de as componentes da série se manifestarem de diferentes formas em diferentes níveis de agregação temporal para com isso modelizar essas componentes em função desses níveis, ou seja, tanto a tendência como o nível são componentes que podem ser modelizados de forma transversal em qualquer nível uma vez que se manifestam ao longo de todos eles, porém, a sazonalidade é uma componente que só pode ser modelizada em níveis de agregação baixos visto que só se manifesta nestes níveis de agregação.

Outro facto observado quando se realiza uma agregação temporal sobre dados intermitentes é que a intermitência e a variância não se mantêm constantes ao longo da variação da frequência. Por exemplo, é simples compreender que ao agregarmos dados a intermitência pode diminuir e eventualmente desaparecer, o que comprova que o grau de intermitência varia nestes casos. Esta situação motiva a que em cada nível de agregação se possa utilizar um método de previsão diferente, uma vez que podemos, por exemplo, deixar de precisar de utilizar métodos de intermitência e poder recorrer a métodos próprios para FMI. Porém, é possível recorrer-se sempre ao mesmo método de previsão em todos os níveis de agregação. Nos trabalhos de Spithourakis et al (2011) e Nikolopoulos et al (2011) ficou provado que a performance média de um único método aplicada a todos os níveis melhora se se recorrer à agregação temporal.

Kourentzes et al. (2014) propõem que a agregação temporal se realize em múltiplas frequências. Ou seja, se tivermos, por exemplo, dados diários transformamo-los em dados semanais, e os dados semanais convertemo-los em mensais e assim sucessivamente. Este procedimento de agregação, também conhecido por agregação temporal sem sobreposição,

foi primeiramente abordado, de uma forma mais superficial, por Willemain et al. (1994) que realizou um estudo numa pequena amostra de séries temporais com incidência em dados intermitentes, com posterior aprofundamento por parte de Nikolopoulos et al. (2011) com um estudo muito mais completo e diversas simulações num conjunto alargado de séries de dados, tendo daí resultado o *Aggregate-Disaggregate Intermittent Demand Approach* (ADIDA) como uma ferramenta de previsão. Neste trabalho os autores abordaram a seleção do nível de agregação apropriado para o conjunto de dados em estudo e um mecanismo de desagregação que permite obter previsões na frequência original. Numa análise do impacto do ADIDA para o meio empresarial, Babai et al. (2012) comprovam que a sua utilização permite aumentar a performance do controlo de *stock* resultando em melhor eficiência de custos e um nível de serviço superior.

O método de agregação temporal de um conjunto de dados levanta de imediato uma questão fulcral para a sua aplicação: qual o nível apropriado de agregação temporal em que devemos parar o nosso processo de agregação?

Esta questão tem sido alvo de discussão existindo alguns estudos que testam diferentes teorias sobre esse limite e apresentam algumas conclusões. Nikolopoulos et al. (2011) sugere, por exemplo, que o nível ideal de agregação para ter uma melhor performance dos métodos seria, sensivelmente, oito períodos deixando em conclusão que a sustentabilidade deste valor exigia ainda continuado estudo. Simultaneamente, no mesmo trabalho, os autores apresentam também a possibilidade de estabelecer um nível de agregação máximo igual ao apresentado na equação (34), que basicamente corresponde à utilização do comprimento do *Lead-Time* (LT) mais um período de revisão (R).

$$\text{Nível de agregação} = LT + R \quad (34)$$

Segundo eles esta heurística apresenta resultados bastante interessantes e permite uma análise mais prática na perspetiva de inventário.

Spitthourakis et al. (2014) desenvolvem um trabalho em torno da ferramenta ADIDA, e conseguem obter resultados interessantes, com melhorias significativas até níveis de agregação em torno de 12 períodos, comprovando, contudo, que, para níveis de agregação muito elevados as previsões obtidas são afetadas por uma suavização excessiva. Enquanto que Rostami-Tabar et al. (2013) demonstram nos seus estudos que a performance dos métodos em obter previsões melhora positivamente com o nível da agregação temporal da série em causa.

Na Figura 8 é apresentado um exemplo de uma agregação temporal com um nível máximo de agregação de 12 níveis.

2.1.6 Intermittent Multiple Aggregation Prediction Algorithm

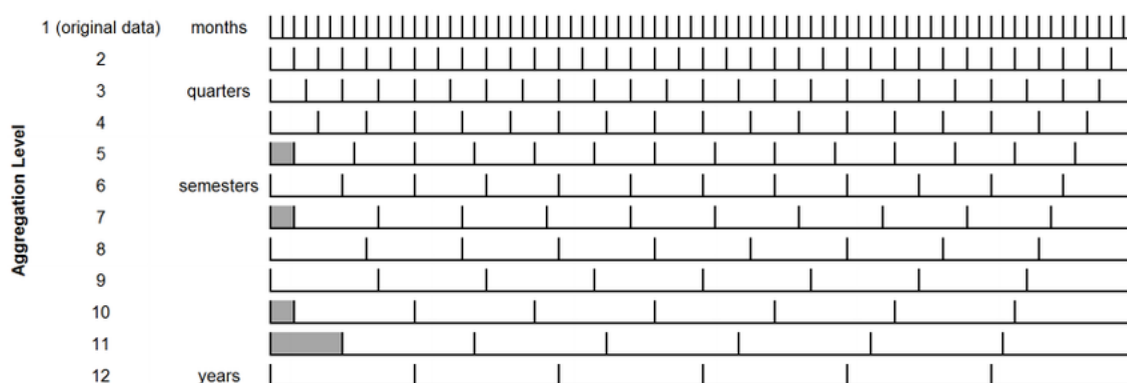


Figura 8 - Agregação temporal sem sobreposição com 12 níveis.

No primeiro nível, identificado como 1 na imagem, encontra-se o conjunto de dados na frequência original e corresponde a valores mensais de consumo. Os dados vão sendo agregados em cada nível através a adição das observações de cada um dos grupos que é agregado. No fundo, pensando no exemplo da passagem do nível 1 para o nível 2, e em relação aos dois primeiros valores de procura, o que acontece é que ambos são somados e o resultado desse cálculo origina um valor único que é o primeiro valor da série agregada do nível dois. Daí que, a não ser nos casos em que o conjunto de dados agregados seja constituído apenas por zeros, a tendência seja a intermitência desaparecer se o conjunto de dados em análise for suficientemente grande. A restante agregação é realizada da mesma forma que foi descrita para o exemplo dos dois primeiros níveis, sendo necessário ter em atenção situações como o nível de agregação cinco onde podemos ver um grupo temporal a sombreado. Isto significa que as observações que estão marcadas a cinza não são tidas em conta quando se realiza a agregação para o nível seguinte, sendo deixadas de fora. Esta truncatura da série tem origem no facto de ser necessário criar grupos com igual tamanho. Por exemplo, no caso do nível cinco da imagem, e sabendo que o conjunto de dados originais possui 72 amostras, temos de remover as duas primeiras observações por forma a conseguirmos construir 14 observações agregadas de cinco valores cada.

Uma solução proposta com o objetivo de solucionar a escolha do nível de agregação ótimo foi introduzida por Petropoulos et al. (2015) e baseia-se na utilização combinada de múltiplos níveis de agregação obtendo com isto uma melhoria da performance de previsão. Assim, o problema de ter que optar por um nível de agregação específico desaparece.

Na Figura 9 é introduzido um exemplo do funcionamento dos quatro passos utilizados para gerar uma previsão através do recurso ao ADIDA. Conforme podemos verificar na imagem, no ponto A temos a série de dados na sua frequência original que apresenta dados mensais de vendas, num total de 24 meses (3 dos quais usados para teste). Na passagem do ponto A para o ponto B é realizada uma agregação de nível 3 transformando os dados mensais em dados trimestrais. O passo seguinte consiste em aplicar o método de previsão selecionado. Conforme foi já referido, o grau de intermitência dos dados não é estático ao longo dos níveis de agregação devendo, numa perspetiva de otimização, realizar-se uma seleção do método mais adequado para aplicar nesse nível. Neste exemplo, foi utilizado o método *Naïve* o que nos permitiu obter no ponto C uma previsão, que é igual ao último valor observado do conjunto de dados agregados que se vêm no ponto B. O conceito de seleção do método mais adequado por níveis vai ser explorado mais à frente. Obtendo uma previsão para o nível de

agregação selecionado existe a necessidade de converter o valor obtido de novo para o nível de agregação de dados original. Este procedimento é referido na literatura como desagregação temporal. A desagregação temporal pode ocorrer de diversas formas sendo a mais amplamente utilizada denominada por *Equal Weights* (EQW) e baseia-se em dividir o valor previsto pelo nível de agregação em que ele se encontra, ou seja, neste caso o nível de agregação é três e a nossa previsão é 15 pelo que obtemos, conforme demonstra a equação (35), três períodos, o 22, 23 e 24, com valores de previsão de consumo de 5 unidades:

$$\text{Valor desagregado} = \frac{\text{previsão}}{\text{nível de agregação}} = \frac{15}{3} = 5 \quad (35)$$

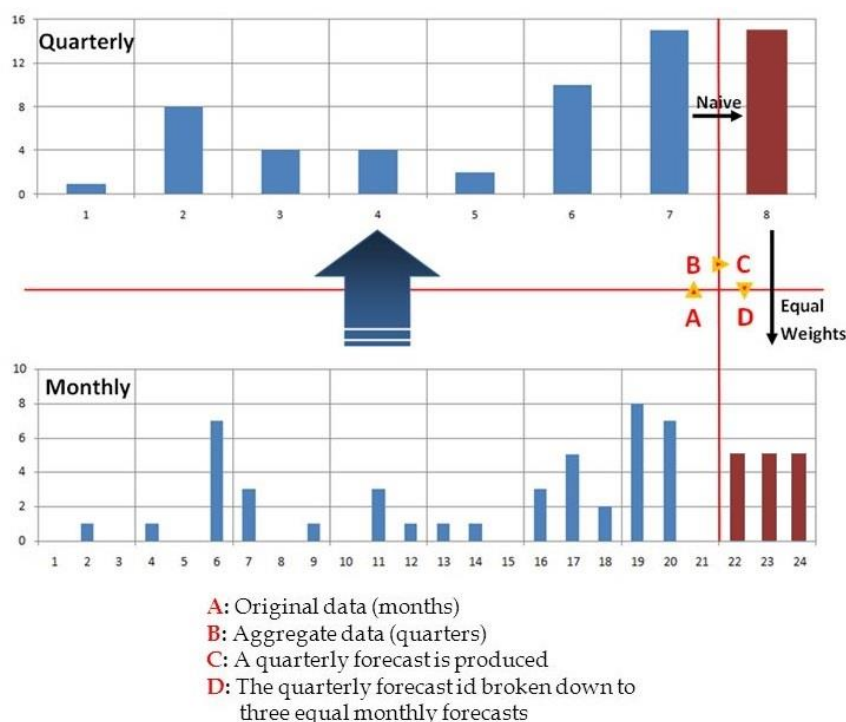


Figura 9 - Previsão com ADIDA (Nikolopoulos et al, 2011).

Para além desta desagregação, baseada na determinação de pesos iguais para os períodos temporais existem mais dois métodos que Nikolopoulos et al. (2011) testam no seu trabalho, mas que, segundo as suas conclusões, têm um desempenho insatisfatório quando submetidos à medida de erro MASE. O primeiro dos métodos é o *PRevious Weights* (PRW), que se baseia na determinação do rácio da observação no grupo temporal mais recente, e o segundo método é o *AVerage Weights* (AVW), e é determinado como a média dos rácios ao longo de todos os grupos temporais não sobrepostos.

A Figura 10 apresenta a análise comparativa realizada por Nikolopoulos et al. (2011) onde é sumariado o erro que eles obtiveram para cada um destes métodos de desagregação temporal numa previsão realizada com o método de *Naïve* e com o SBA num total de 24 níveis de agregação temporal diferentes. Do lado esquerdo da Figura 10 encontram-se os resultados obtidos pela utilização do método *Naïve*. Nesta representação a primeira linha, de cima para baixo, a bordô corresponde às previsões desagregadas com recurso ao PRW, a segunda que se encontra a verde corresponde às previsões obtidas pela desagregação recorrendo ao AVW, e

por fim a linha a azul corresponde às previsões obtidas com desagregação por intermédio do EQW. Uma vez que a medida de erro em utilização é o MASE, e analisando as linhas registadas neste gráfico é de conclusão direta que o EQW apresenta melhor desempenho ao longo de todos os níveis de agregação testados, uma vez que possui um valor de MASE mais baixo que os restantes.

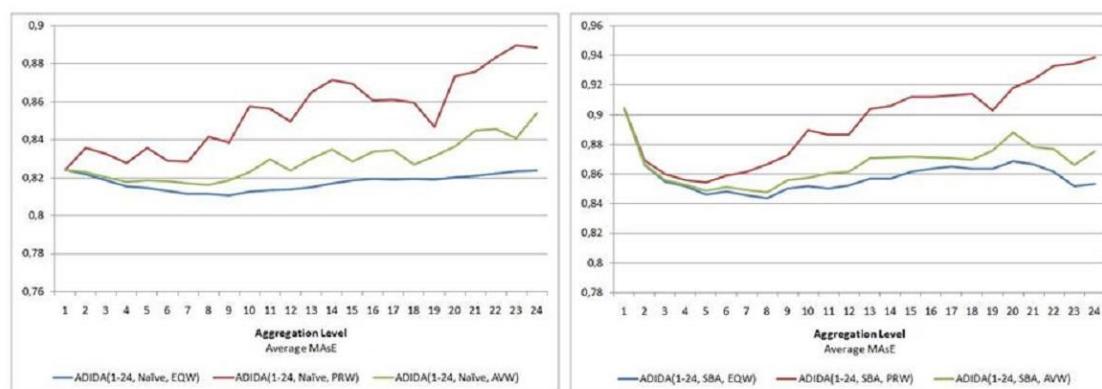


Figura 10 - MASE dos níveis de agregação para os vários métodos de previsão/desagregação (Nikolopoulos et al., 2011)

No gráfico que visualizamos do lado direito são apresentados os resultados obtidos para previsões obtidas com recurso ao SBA. A linha apresentada na imagem mantém a mesma distribuição, ou seja, bordô corresponde ao PRW, verde ao AVW e azul ao EQW e a medida de erro utilizada mantém-se o MASE. Conforme podemos comprovar no gráfico a superioridade do EQW permanece ao longo de todos os 24 níveis de agregação, manifestando uma superação desse método em comparação com os restantes dois à medida que os níveis de agregação aumentam. Nikolopoulos et al. (2011) replicaram esta experimentação para poder avaliar o comportamento nestas mesmas condições, mas recorrendo a outras duas medidas de erro por forma a conseguir sustentar as conclusões que retiraram da análise que aqui é apresentada. Após realizarem esta simulação avaliando o erro com o MSE (*Mean Square Error*) e o GMRMSE (*Geometric Mean Relative Mean Square Error*) a conclusão a que chegaram com os resultados obtidos é a de que o EQW obteve uma performance superior aos restantes dois métodos em todas as simulações realizadas.

A eficiência da metodologia de agregação temporal foi profundamente estudada por Spithourakis et al. (2011) que testaram o seu desempenho em métodos de previsão amplamente estudados e aceites, demonstrando melhorias muito significativas em métodos como o SES, Theta e *Naïve* (Assimakopoulos et al., 2000).

O objetivo ao recorrer a diferentes níveis de agregação prende-se com o facto de que diferentes níveis apresentam, ou reforçam, determinadas características do conjunto de dados, por exemplo, quanto mais elevado for o nível de agregação maior foco é obtido nas componentes de baixa frequência como a tendência ou o nível, enquanto que quanto menos agregada a série estiver maior destaque é dado às componentes de alta frequência como a sazonalidade. Com a agregação quando umas componentes são atenuadas as outras são destacadas.

A agregação temporal é um processo que ao combinar os dados em blocos leva a uma diminuição do volume de dados disponíveis, o que pode eventualmente ser um problema principalmente se o conjunto em análise não possuir uma dimensão suficiente. Porém, a possibilidade de eliminação, ou redução, dos zeros existentes ao realizar esta conjugação em blocos é bastante interessante para a aplicação dos métodos de previsão e consequente performance das previsões realizadas.

Em suma, quer o iMAPA quer o ADIDA baseiam-se em processos de agregação e desagregação de conjuntos de dados de séries temporais para realizar previsões para períodos futuros. A fundamental diferença entre estes dois métodos é que enquanto o ADIDA utiliza apenas um único nível de agregação temporal pré definido, aplica o método de previsão pretendido e depois desagrega o resultado obtido, o iMAPA executa múltiplas agregações temporais de dados, gerando em simultâneo um conjunto composto por dados de vários níveis de agregação diferentes com o objetivo de explorar as vantagens de cada série temporal agregada e original, sendo que, não existindo um nível de agregação pré-definido, são geradas várias previsões para o mesmo item, sendo que as previsões obtidas necessitam ser combinadas num valor de previsão único que será a nossa previsão final deste método. De certa forma, de uma maneira muito simplista, e em termos de conceito de aplicação, podemos pensar no iMAPA como um conjunto de ADIDAS realizados em simultâneo e cujas previsões obtidas são posteriormente combinadas para gerar um *output* único.

Este sistema de combinação de previsões tem várias abordagens possíveis que têm sido foco de reflexão da literatura nos últimos anos. Primeiramente, esta metodologia de combinação de previsões tem sido bem aceite pelos analistas que se debruçam sobre ela como Bates et al. (1969), Makridakis et al. (1983) ou Clemen (1989) que consideram que a sua utilização traz benefícios para a previsão gerada pela combinação dos resultados obtidos através de níveis diferentes. As previsões obtidas pela combinação de previsões têm provado ser mais robustas e geram erros menores, ou seja, obtêm-se previsões mais próximas da procura real manifestada. Este facto deve-se a um melhoramento da variância e da incerteza nas previsões geradas (Hibon e Evgeniou, 2005). Existem diversas abordagens possíveis em termos de seleção de metodologia de conjugação das previsões obtidas, sendo que maioritariamente os analistas optam por metodologias simples que permitam uma implementação direta e pouco complexa. Porém têm surgido recentemente estudos no sentido de desenvolver abordagens mais complexas das quais se possa retirar uma performance superior em termos de previsão obtida. Investigadores como o Kolassa (2011), Taylor (2008) ou He et al. (2005) desenvolvem pesquisa no sentido de obter sistemas sofisticados de seleção de pesos para a combinação de previsões existindo, contudo, uma ampla sustentação e defesa da utilização de métodos com muito menor grau de complexidade, como médias que recorram à utilização de igualdade de pesos para as diversas componentes, e que se têm provado ter um desempenho robusto em diversos estudos quando comparados com metodologias mais complexas (Jose et al., 2008; Hibon et al., 2005; Clemen, 1989; Timmermann, 2006).

Se analisarmos os estudos que têm sido realizados acerca da combinação de previsões a grande maioria encontra-se centrada em combinar previsões pontuais obtidas de metodologias de previsão diferentes a partir do mesmo conjunto de dados, existindo, porém, uma outra abordagem bastante interessante que recorre à combinação de previsões obtidas a diferentes frequências a partir do mesmo conjunto de dados. É esta abordagem que o iMAPA

implementa e lhe permite captar e conjugar as diferentes dinâmicas dos dados manifestadas a diferentes níveis.

Cholette (1982) é o primeiro a apresentar desenvolvimento científico no sentido da combinação de previsões obtidas a várias frequências, realizando uma alteração às previsões mensais obtidas pelo modelo ARIMA através da utilização de previsões obtidas pelo julgamento de analistas que utilizaram dados de frequências alternativas. Trabelsi et al. (1989) mantêm o estudo focado no modelo ARIMA mas analisando já a combinação de previsões obtidas por recurso a séries de diferentes frequências temporais.

Neste ponto da secção que aborda o iMAPA já foram abordados praticamente todos os tópicos relevantes que explicam e sustentam o seu funcionamento à exceção de um tópico que se apresenta de significativa importância, sendo este, a seleção do método a aplicar na previsão realizada em cada nível de agregação dos dados.

Conforme foi referido atrás, quando foi introduzido o exemplo da Figura 9, a seleção do método para realizar a previsão nesse caso foi baseada no julgamento do próprio analista, ou seja, ele estabeleceu o método que achou mais adequado.

Porém, se no caso do método ADIDA esta escolha é predefinida, no caso do iMAPA, onde temos múltiplos níveis de agregação diferentes com diferentes características, a escolha de um único método de previsão transversal a todos os níveis gerados poderá ser uma simplificação grosseira.

Tabela 9 - Exemplo da variação do grau de intermitência.

	Níveis				
Período	1	2	3	4	5
0					
1	1,36449919				4,329093683
2	0	0	1,364499190		
3	0				
4	2,964594493	2,964594493	3,627994710	10,830131008	
5	0				
6	0,663400217	7,865536515		10,830131008	15,524685140
7	7,202136298				
8	3,533886847	5,832821957	13,034958255	7,659148626	
9	2,29893511				
10	1,826326669	1,826326669	1,826326669	7,659148626	2,604258817
11	0				
12	0	0,356198355		2,604258817	
13	0,356198355				
14	0	2,248060461	2,604258817	2,604258817	
15	2,248060461				

Na tabela 9 é apresentado de uma forma muito simplista o que aconteceria à nossa série de dados se lhe aplicássemos vários níveis de agregação temporal.

É de notar que a série não é suficientemente comprida para esta metodologia ser utilizada de forma efetiva, sendo que neste caso o objetivo é apenas o de mostrar que à medida que vamos agregando e considerando níveis de agregação superiores vai ocorrendo uma alteração do grau de intermitência dos dados em análise podendo eventualmente levar a um desaparecimento total da intermitência.

Tendo esta noção de que a intermitência varia em grau ao longo dos níveis de agregação e que pode inclusive desaparecer, impõem-se uma reflexão acerca da aplicação de um método único a todos os níveis de agregação. Neste sentido, torna-se lógico que o método utilizado em cada nível de agregação em conjuntos de dados intermitentes deva ser selecionado em cada nível de agregação por forma a este seja o método mais adequado às características dos dados naquele nível. É aqui que se torna relevante o recurso a métodos de classificação que permitam a seleção dos métodos de previsão mais adequados.

Existem três esquemas de classificação para séries de dados intermitentes, sendo que um dos mais abordados na literatura é o apresentado por Syntetos et al. (2005). Numa primeira fase a série é categorizada como “suave”, “intermitente”, “errática” e “irregular” e de acordo com essa categorização opta-se pela aplicação do SBA ou do CRO. A categorização da série é feita de acordo com o valor de dois parâmetros: (1) o p , que é o valor médio do intervalo entre procuras não nulas, e (2) o v , que é o quadrado do coeficiente de variação da procura. Segundo estes autores, quando a série é categorizada como “suave” o método a utilizar é o CRO, e nos restantes casos deverá recorrer-se ao SBA, conforme se mostra na Figura 11.

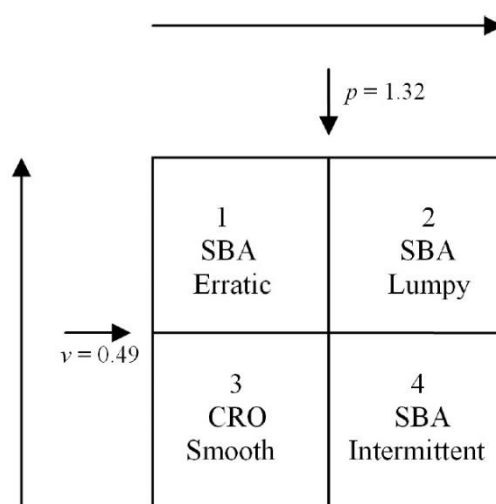


Figura 11 - Esquema de categorização SBC (Kostenko, 2006).

Conforme se identifica na imagem, os valores limite que definem os quadrantes são $v = 0.49$ e $p = 1.32$. Este sistema de categorização é baseado no erro quadrático médio de cada método de previsão e, assumindo dados identicamente distribuídos e independentes, o MSE do CRO é superior ao do SBA se:

$$v > \frac{4p(2-p)-\alpha(4-\alpha)-p(p-1)(4-\alpha)(2-\alpha)}{p(4-\alpha)(2p-\alpha)} \quad (36)$$

Esta é a equação que define a área de seleção dos métodos e que é representada na Figura 12, onde estão representados limites de separação entre os dois métodos para diferentes valores de α (alfa).

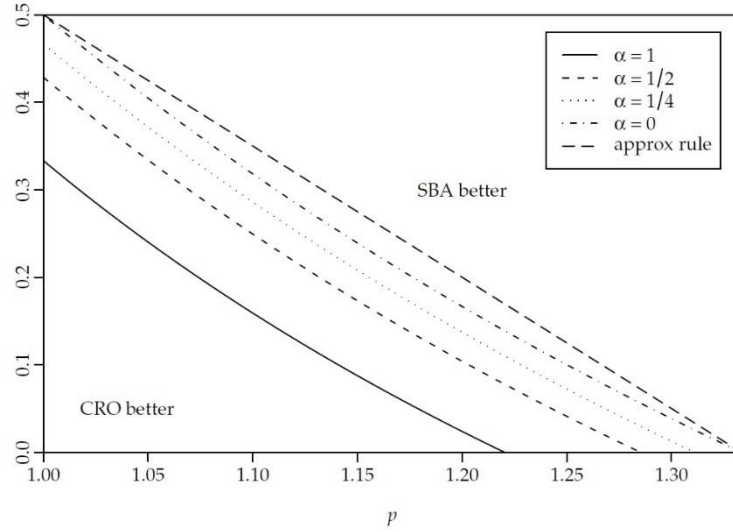


Figura 12 - Espaço de parâmetros para os quais um método tem melhor desempenho que o outro.

Kostenko et al. (2006) sugerem uma regra mais simples de divisão do espaço de parâmetros para facilitar a interpretação do método selecionado. Eles sugerem que a regra de definição dos limites passe a ser a apresentada na equação (37) que de uma forma visual significa estabelecer como linha de separação dos dois métodos a linha que divide o quadrante diagonalmente em dois, que na Figura 12 corresponde à linha assinalada como “approx rule”.

$$v > 2 - \frac{3}{2}p \quad (37)$$

Heinecke et al. (2013) comprovam, num estudo realizado sobre o comportamento do esquema de classificação proposto por Syntetos et al. em comparação com a alteração proposta por Kostenko et al., que a proposta revista deste método apresenta uma performance de previsão globalmente melhor.

Mais tarde, Petropoulos et al. (2015) dão continuidade ao estudo deste esquema de classificação criado por Syntetos, Boylan e Croston, e melhorado por Kostenko e Hyndman, e sugerem uma nova versão deste esquema de classificação motivados pelo facto de no aumento dos níveis de agregação de dados existir a possibilidade de a intermitência desaparecer por completo. O que eles propõem é que no caso de deixar de existir intermitência no conjunto de dados, que é equivalente a dizer que $p = 1$, se deve passar a utilizar como método de previsão o SES, em vez de métodos que foram projetados para previsão em conjuntos de dados intermitentes, e ainda recorrer a utilização de um parâmetro de alisamento otimizado para o nível do SES em vez de um parâmetro de alisamento estático conforme é sugerido para o CRO ou o SBA. Demonstram ainda que a utilização do CRO, nestes casos, não é equivalente ao SES pois apesar de todos os períodos terem procura diferente de zero essa equivalência de métodos não se verifica para coeficientes de variação superiores,

onde o SBA seria escolhido neste esquema anteriormente proposto, justificando assim a pertinência da atualização que propuseram a este esquema de classificação.

Com a abordagem dos sistemas de classificação para seleção de métodos de previsão, foram já exploradas todas as componentes que definem a metodologia iMAPA, que se apresenta como uma das mais complexas das que foram utilizadas no desenvolvimento deste trabalho. Em suma, o iMAPA objetiva tirar partido de algumas características e conjuga-as de forma a beneficiar de melhores performances nas previsões que realiza. Para o efeito recorre à agregação temporal, à seleção de métodos adequados por recurso a esquemas de classificação, à desagregação e à combinação de previsões.

2.2. Modelos de espaço de estados

Até à presente secção foi apresentado o percurso da evolução dos diversos métodos de previsão de procura intermitente, da forma como eles foram desenvolvidos, estruturados e apresentados pelos seus autores.

De seguida vai ser introduzida uma outra forma de abordar estes métodos recorrendo-se ao conceito de modelo de espaço de estados que é constituído por duas partes: uma parte mensurável e uma parte inobservável. A primeira corresponde à parte observável constituída pelos componentes, e a segunda corresponde à evolução dessas componentes.

Por forma a tornar este conceito mais explícito Svetunkov (2017) introduz o exemplo seguinte. Pensando numa série que não apresenta nem tendência nem sazonalidade, o seu modelo de espaço de estados é descrito por:

$$\begin{aligned} y_t &= l_{t-1} + \epsilon_t \\ l_t &= l_{t-1} + \alpha \epsilon_t \end{aligned} \quad (38)$$

Onde y_t como o valor da série temporal no instante t , l_t o valor da componente do nível no instante t , $\epsilon_t \sim NID(0, \sigma^2)$ é o erro e α (alfa) é um parâmetro.

Com o desenvolvimento do seu trabalho e foco nos modelos de espaços de estados Svetunkov (2017) propõem uma regra que afirma ter características extremamente interessantes em termos de aplicação na previsão, que possibilita a representação de praticamente qualquer processo estatístico por recurso a modelos de espaço de estados.

Svetunkov baseia a implementação de todos os modelos que disponibiliza no seu package a partir de uma forma geral de espaço de estados com fonte única de erro que é apresentada na equação (39) e que é similar à apresentada por Hyndman et al. (2008):

$$\begin{aligned} y_t &= o_t(w(v_{t-l}) + r(v_{t-l})\epsilon_t, \\ v_t &= f(v_{t-l}) + g(v_{t-l})\epsilon_t \end{aligned} \quad (39)$$

Neste modelo geral y_t corresponde a previsão do momento presente, o_t é modelado por uma distribuição de Bernoulli, ou seja, é uma variável binária que é 1 quando y_t é observado o que confere bastante flexibilidade ao modelo permitindo que este sirva para os casos de *fast-moving goods* e *slow-moving goods* possibilitando analisar séries de dados intermitentes; ϵ_t é

o termo do erro; v_{t-l} é um vetor de estado que apresenta a estrutura apresentada na equação (40) e é composta por L , que é uma matriz de atrasos presente na equação (41), composta por operadores de mudança de turno, B , e componentes de atraso, m_k , sendo que, por exemplo, num modelo ETS(A,A,A) (ver anexo 1 a 5) o $B^{m1} = B$ significa que a componente de nível tem atraso 1, $B^{m2} = B$ significa que a componente de tendência também tem atraso 1 e $B^{m3} = B^{12}$ significa que a componente sazonal tem atraso 12; multiplicando o vetor de estados, v'_t , que é composto pelo nível, tendência e sazonalidade e apresentado na equação (42). Ainda neste modelo, equação (39), temos o $w(\cdot)$ que é uma função de medição; $r(\cdot)$ é função de termo de erro; $f(\cdot)$ é uma função de transição; $g(\cdot)$ é uma função de persistência e l representa um vetor de atrasos, sendo que as funções $w(\cdot)$, $r(\cdot)$, $f(\cdot)$ e $g(\cdot)$ permitem a oscilação do modelo de espaço de estados entre componentes aditivas e multiplicativas.

O modelo ETS (*Exponential Trend Seasonal*) define um tipo de tendência e sazonalidade específicas em análise. Conforme podemos ver no anexo A.1 a tendência pode ser Nenhuma (N), Aditiva (A), Aditiva amortecida (A_d), Multiplicativa (M) ou Multiplicativa amortecida (M_d); quanto à sazonalidade esta pode ser Nenhuma (N), Aditiva (A) ou Multiplicativa (M). Conforme podemos ver no anexo A.2 podemos classificar os modelos ETS quanto à conjugação efetuada com as diversas possibilidades de componentes de tendência e sazonalidade. As conjugações realizadas com as componentes permite-nos obter diferentes métodos de previsão, com consequentes características diferentes.

Se inserirmos o resultado obtido na equação (40) num modelo de espaço de estados puramente aditivo obtemos o sistema de equações para o modelo em análise e que está representado na equação (43) e que é constituída pela previsão, determinação do nível, tendência e sazonalidade.

A primeira equação apresentada na equação (39) é conhecida por equação de medição ou de observação, sugerindo que os dados são quantificados por recurso a esta equação, enquanto que a segunda equação aí apresentada é denominada equação de transição, o que sugere realizar a transição entre componentes dos modelos de espaço de estados.

$$v_{t-l} = Lv_t = \begin{pmatrix} B^{m1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & B^{mk} \end{pmatrix} v_t \quad (40)$$

$$L = \begin{pmatrix} B^{m1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & B^{mk} \end{pmatrix} \quad (41)$$

$$v'_t = (l_t \ b_t \ s_t) \quad (42)$$

$$\begin{aligned} y_t &= l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \\ l_t &= l_{t-1} + b_{t-1} + \alpha \epsilon_t \\ b_t &= b_{t-1} + \beta \epsilon_t \\ s_t &= s_{t-m} + \gamma \epsilon_t \end{aligned} \quad (43)$$

A partir do modelo geral é possível gerar qualquer um dos diversos tipos de modelos de espaços de estados que Svetunkov utiliza no seu package.

Os modelos desenvolvidos podem ser: aditivos puros, multiplicativos puros e mistos. Pela forma como este package foi construído o modelo ETS, implementado na função `es`, recorre a modelos não aditivos, porém todas as restantes funções do package utilizam apenas modelos aditivos puros.

Este pacote está implementado de forma a que antes da realização de uma previsão, ou conjunto de previsões, exista uma inicialização de variáveis e parâmetros por recurso a processos de otimização. Este processo ocorre para momentos $t < 1$, ou seja, estes modelos são inicializados antes da série temporal começar para que seja possível desde o primeiro valor da série conseguir atualizar as componentes, por recurso à equação de transição, e utilizar a equação de medição para obter previsões com um horizonte temporal de uma unidade.

Para o processo de otimização é necessário realizar a seleção do método de inicialização, dos parâmetros do espaço de restrições e ainda qual a função de custo a utilizar, sendo que estão implementadas neste package as seguintes:

- MSE (Mean Squared Error), obtido pela equação:

$$MSE = \frac{1}{T} \sum_{t=1}^T e_{t+1|t}^2 \quad (44)$$

- MAE (Mean Abslue Error), apresentado na equação:

$$MAE = \frac{1}{T} \sum_{t=1}^T |e_{t+1|t}| \quad (45)$$

- HAM (Half Absolute Moment), calculado por:

$$HAM = \frac{1}{T} \sum_{t=1}^T \sqrt{|e_{t+1|t}|} \quad (46)$$

- MSEh (Mean Squared h steps ahead Error), determinado por:

$$MSE_h = \frac{1}{T} \sum_{t=1}^T e_{t+h|t}^2 \quad (47)$$

- MSTFE (Mean Squared Trace Forecast Error), calculado por:

$$MSTFE = \sum_{j=1}^h \frac{1}{T} \sum_{t=1}^T e_{t+j|t}^2 \quad (48)$$

- MLSTFE (Mean Logarithmic Squared Trace Forecast Error), obtido por:

$$MSTFE = \sum_{j=1}^h \log \left(\frac{1}{T} \sum_{t=1}^T e_{t+j|t}^2 \right) \quad (49)$$

Onde e corresponde à determinação dos resíduos, ou erros, e são calculados para modelos com erros aditivos segundo:

$$e_{t+1|t} = y_{t+1} - \mu_{t+1|t} \quad (50)$$

E para modelos com erros multiplicativos:

$$\tilde{\epsilon}_{t+1|t} = \frac{y_{t+1} - \mu_{t+1|t}}{\mu_{t+1|t}} \quad (51)$$

Sendo $\mu_{t+1|t}$ a previsão da procura no momento $t+1$.

O processo de otimização implementado segue dois passos específicos, primeiro realiza-se uma otimização dos parâmetros aplicando um algoritmo *BOBYQA*, com um máximo de 1000 avaliações e uma tolerância relativa de $1e-8$, e em segundo lugar os parâmetros são ainda mais otimizados por recurso ao algoritmo *Nelder-Mead*, com um máximo de 1000 avaliações e uma tolerância relativa de $1e-6$. Este processo de otimização em dois passos dá-nos boas previsões de parâmetros.

Para entendermos de forma simples como funcionam estes algoritmos passamos a sua análise, começando com o *BOBYQA* que permite otimizar uma função objetivo através da utilização do método de regiões de confiança que cria modelos quadráticos por interpolação; e o algoritmo *Nelder-Mead* otimiza funções objetivo realizando uma pesquisa heurística que converge para pontos não estacionários.

Segundo Snyder (2002) e Hyndman et al (2006) os modelos estatísticos para a análise de dados intermitentes devem ser modelos multiplicativos.

Neste pacote, todos os modelos desenvolvidos para dados intermitentes utilizam a letra “i” por forma a referenciar que são métodos específicos para este tipo de dados, sendo que as propriedades abordadas são válidas para dados com e sem intermitência, contudo é de referir que existem alterações essencialmente em termos de obtenção da expectativa condicional e da variância necessárias à determinação da probabilidade de ocorrência futura.

2.2.1. Modelo de espaço de estados puramente aditivo para intermitência

O modelo de espaço de estados representado de forma puramente aditiva é o modelo que é demonstrado na equação (52).

$$\begin{aligned} y_t &= o_t(w'v_{t-l} + \epsilon_t) \\ v_t &= Fv_{t-l} + g\epsilon_t \end{aligned} \quad (52)$$

Sendo $\epsilon_t \sim N(0, \sigma^2)$.

Este modelo é específico para conjuntos de dados que possuam valores positivos e negativos nas suas observações, sendo este o facto que explica que este modelo de espaço de estados não seja o modelo ideal a aplicar a séries temporais de procura, por exemplo, de produtos de um retalhista, uma vez que essas séries não possuem, pelas suas características, valores negativos. De forma similar ao modelo apresentado anteriormente, o_t continua a ser uma variável que pode ser aproximadamente modelada por uma distribuição de Bernoulli(p_t).

A equação de medição correspondente ao método pode ser descrita conforme a equação (53).

$$y_{t+h} = o_{t+h}(w'F^{h-1}v_t + \sum_{j=1}^{h-1} w'F^{j-1}g\epsilon_{t+h-jl} + \epsilon_{t+h}) \quad (53)$$

Este modelo apresenta ainda uma expectativa condicional de probabilidade e uma variância condicional visíveis na equação (54).

$$\begin{aligned} \mu_{t+h|t} &= \mu_{p,t+h|t} (w' F^{h-1} v_t) \\ \sigma_{t+h|t}^2 &= \sigma^2 \left(1 + \sum_{j=1}^{h-1} (w' F^{j-1} g)^2 \right) (\sigma_{p,t+h|t}^2 + \mu_{p,t+h|t}^2) + (w' F^{h-1} g)^2 \sigma_{p,t+h|t}^2 \end{aligned} \quad (54)$$

O modelo de espaço de estados puramente aditivo assume que existe uma evolução dos estados do modelo, que podem ser vistos como a vontade dos clientes em adquirir um produto, inclusive quando não existe ocorrência de procura, ou seja, quando registamos valores de procura zero

O intervalo de previsão correspondente a este método é calculado com recurso à equação (55).

$$\mu_{t+h|t} + Z_{\frac{\alpha}{2}} \sigma_{t+h|t} < y_{t+h} < \mu_{t+h|t} + Z_{1-\frac{\alpha}{2}} \sigma_{t+h|t} \quad (55)$$

2.2.2. Modelo espaço de estados puramente multiplicativo para intermitência

O modelo de espaço de estados representado de forma puramente multiplicativa é o modelo demonstrado na equação (56).

$$\begin{aligned} y_t &= o_t \exp(w' \log(v_{t-l}) + \log(1 + \epsilon_t)) \\ v_t &= \exp(F \log(v_{t-l}) + \log(1 + g \epsilon_t)) \end{aligned} \quad (56)$$

Sendo $(1 + \epsilon_t) \sim \log N(0, \sigma^2)$.

Este modelo é considerado o modelo ideal para ser aplicado aos casos de dados intermitentes quando comparado com os restantes modelos de espaços de estados. A variável o_t continua a poder ser aproximadamente modelada por uma distribuição de Bernoulli(p_t).

A equação de medição obtida para este modelo é apresentada na equação (57).

$$y_{t+h} = o_t \exp(w' \log(v_t) + \sum_{j=1}^{h-1} w' F^{j-1} \log(1 + g \epsilon_{t+h-jl}) + \log(1 + \epsilon_{t+h})) \quad (57)$$

O modelo possui uma expectativa condicional de probabilidade e uma variância condicional definidas na equação (58).

$$\begin{aligned} \mu_{t+h|t} &= \mu_{p,t+h|t} \exp\left(\tilde{\mu}_{t+h|t} + \frac{\tilde{\sigma}_{t+h|t}^2}{2}\right) \\ \sigma_{t+h|t}^2 &= \hat{\sigma}_{t+h|t}^2 (\sigma_{p,t+h|t}^2 + \mu_{p,t+h|t}^2) + \sigma_{p,t+h|t}^2 \end{aligned} \quad (58)$$

É determinado, ainda, um intervalo de previsão, que recorre à primeira e segunda equações da equação (54), e que está demonstrado na equação (59).

$$\mu_{t+h|t} \left(1 - q_{h, \frac{\alpha}{2}} \sigma_{t+h|t}\right) < y_{t+h} < \mu_{t+h|t} \left(1 + q_{h, 1-\frac{\alpha}{2}} \sigma_{t+h|t}\right) \quad (59)$$

2.2.3. Croston

O modelo de Croston implementado no package *Smooth*, disponível no CRAN e utilizado nas análises realizadas com recurso ao R *studio*, segue o modelo original proposto por Croston (1972), conforme a equação (60), e a análise feita por Hyndman et al. (2008) que demonstra a existência de dois modelos estatísticos ETS(A,N,N), modelo de alisamento exponencial simples com erros aditivos, e ETS(M,N,N), modelo de alisamento exponencial simples com erros multiplicativos, conforme demonstra nomenclatura dos anexos 1 a 5 e é apresentado em Hyndman et al (2008).

$$p_t = \frac{1}{q_t} \quad (60)$$

Nesta equação q_t corresponde ao valor de procuras nulas consecutivas entre observações e a probabilidade de ocorrência, p_t , é atualizado em cada momento de procura registada e diferente de zero, e o seu valor diminui à medida que o número de zeros aumenta na amostra, já que com o aumento do número de zero existe um aumento equivalente de q_t .

Este método é estruturado conforme exposto na equação (61).

$$\begin{aligned} q_t &= l_{q,t-1}(1 + \epsilon_t) \\ l_{q,t} &= l_{q,t}(1 + \alpha_q \epsilon_t) \end{aligned} \quad (61)$$

Onde $(1 + \epsilon_t) \sim \log N(0, \sigma^2)$.

Possui uma expectativa condicional de probabilidade, cuja demonstração do cálculo foi realizada por Syntetos et al. (2005), e uma variância condicional determinadas na equação (62).

$$\begin{aligned} \mu_{p,t+h|t} &= \left(1 - \frac{\alpha_q}{2}\right) \frac{1}{\mu_{q,t+h|t}} \\ \sigma_{p,t+h|t}^2 &= \left(1 - \frac{\alpha_q}{2}\right) \frac{\mu_{q,t+h|t} - \left(1 - \frac{\alpha_q}{2}\right)}{\mu_{q,t+h|t}^2} \end{aligned} \quad (62)$$

2.2.4. Teunter, Syntetos e Babai

O mdelo TSB baseia-se no trabalho de Teunter et al. (2011), sendo a sua implementação baseada no modelo ETS(M,N,N) que Svetunkov (2017) considera mais adequado para a implementação deste método do que realizar a estimação da probabilidade e atualização através do SES, obtendo-se neste caso o método exposto na equação (63).

$$p'_t = l_{p,t-1}(1 + \epsilon_t) \quad (63)$$

$$l_{p,t} = l_{p,t-1}(1 + \alpha_p \epsilon_t)$$

Com $p'_t = (1 - 2k)p_t + k$ e k é um número muito pequeno, implementado no smooth com o valor $k = 10^{-5}$. Assim, podemos afirmar que p'_t pode ser aproximadamente modelado por uma distribuição Beta(a,b) e pode recorrer-se a inferência estatística.

Esta implementação realiza uma alteração à estimação de probabilidade, p_t , que está associada com o facto de a sua distribuição, log-normal, só ser limitada de por um lado podendo gerar valores de probabilidade superiores a 1, o que não teria qualquer lógica quando interpretado. Pelo que a solução passa por recorrer a uma distribuição que não permita que este tipo de situação ocorra, sendo sugerida por Svetunkov a distribuição Beta e a adaptação que se pode ver na equação (63), uma vez que esta distribuição é flexível permitindo abranger diferentes distribuições possíveis, fazendo referência ao facto que a distribuição Beta não permite que os valores fronteira sejam assumidos.

Este modelo no formato de espaço de estados possui uma expectativa condicional de probabilidade e uma variância condicional determinada na equação (64).

$$\begin{aligned} \mu_{p,t+h|t} &= l_t \\ \sigma_{p,t+h|t}^2 &= \sigma_p^2(1 + (h-1)\alpha_p^2) \end{aligned} \tag{64}$$

Teunter et al. (2011) propõem a utilização do método de Naïve para a estimação de p_t , ou seja, se existir ocorrência de procura então $p_t = 1$, nos restantes casos é zero. Svetunkov refere a dificuldade de implementar esta característica através do recurso à inferência, mas, uma vez que esta particularidade confere ao método a importante capacidade de atualização da probabilidade quando a procura é zero, é imperativo a manutenção desta característica.

Svetunkov, no pacote *Smooth* que desenvolveu e ao qual recorreremos na fase de simulação, realiza a implementação deste método através de dois passos distintos. Numa primeira fase ele estima os parâmetros da distribuição Beta(a,b) e o parâmetro α (alfa), sendo que numa segunda fase são estimados os parâmetros da equação (63).

2.2.5. Fixed e Auto

Em relação à função Fixed que utilizamos na análise do nosso conjunto de dados é simples de entender o seu funcionamento, e o que esta faz é possibilitar ao analista definir uma probabilidade fixa de ocorrências que o modelo assume e que não varia com o tempo; Svetunkov desenvolve e disponibiliza ainda a função “Auto” que não existe no tsintermittent e que no seu conceito apresenta-se interessante enquanto metodologia de análise e despertou a curiosidade na sua utilização.

Assim, analisa-se em particular a função Auto para perceber melhor o seu funcionamento. Esta é uma função que no parâmetro “intermittent” permite selecionar várias opções de entre vários métodos a serem aplicados à série temporal, sendo que

uma das opções disponíveis é o “Auto”, e esta é uma opção bastante interessante. O que o “Auto” faz é analisar a série temporal com recurso ao critério de informação AIC que é determinado segundo a equação (65), que utiliza a determinação das funções de semelhança para os modelos de espaço de estados que são apresentadas na equação (66), para erros aditivos, e na equação (67), para erros multiplicativos, e em função dos resultados obtidos no critério de informação seleciona e aplica automaticamente o modelo de previsão que obteve o menor valor de critério de informação já que este é aquele que melhor se adapta à série temporal em análise.

$$AIC = 2k - 2l(\theta, \hat{\sigma}^2|Y) \quad (65)$$

Com k a ser o número de parâmetros e $l(\theta, \hat{\sigma}^2|Y)$ a ser o logaritmo da verosimilhança que no caso do erro ser aditivo é obtido pela equação 66.

$$l(\theta, \hat{\sigma}^2|Y) = -\frac{T_1}{2}(\log(2\pi e) + \log(\hat{\sigma}^2)) + \sum_{o_t=1} \log(p_t) + \sum_{o_t=0} \log(1 - p_t) \quad (66)$$

Com Y a ser um vetor com todos os valores de procura e T_1 a ser o número de observações diferentes de zero.

No caso dos erros serem multiplicativos a verosimilhança é obtida pela equação 67.

$$l(\theta, \hat{\sigma}^2|Y) = -\frac{T_1}{2}(\log(2\pi e) + \log(\hat{\sigma}^2)) - \sum_{o_t=1} \log y_t + \sum_{o_t=1} \log(p_t) + \sum_{o_t=0} \log(1 - p_t) \quad (67)$$

Um dos parâmetros que foi tido em atenção, e que o Svetunkov reforça, é que para se utilizar esta função deve-se recorrer ao MSE como função de custo, pois se não o fizermos ele vai ser calculado, mas não otimizado, e posteriormente utilizado para determinar a função de semelhança e o critério de informação que por consequência não estarão otimizados podendo levar à ocorrência de erros.

O *Smooth* apresenta um conceito de implementação interessante pelo que apresenta-se como uma escolha interessante para a análise da nossa série de dados, possibilitando ainda a comparação do seu desempenho com a performance obtida da análise realizada com recurso ao *tsintermittent*.

Capítulo 3

Caso de Estudo

3.1. Caso Jerónimo Martins

O presente capítulo aborda três pontos fulcrais associados com trabalho levado a cabo.

Primeiramente, é feita uma análise do conjunto de dados que foram gentilmente fornecidos pelo Grupo Jerónimo Martins e que nos permitiu a realização deste trabalho de estudo e análise de procura intermitente.

Seguidamente, é abordada a metodologia utilizada para realizar o processo de aplicação dos métodos aos dados em estudo bem como as particularidades da ferramenta de desenvolvimento e respetivos pacotes de métodos.

Por fim, é explorada a metodologia utilizada por forma a aferir a performance resultante do trabalho realizado.

3.2. Conjunto de dados

O conjunto de dados que nos possibilitou realizar este trabalho foi fornecido pela Jerónimo Martins ao CESE e é constituído por um conjunto de dados diários referentes às vendas por categoria e por SKU registadas em 12 lojas da marca Pingo Doce que pertence a esse grupo.

É de realçar a extrema importância que o acesso a um conjunto de dados como este que nos foi fornecido traz para o desenvolvimento do trabalho levado a cabo na área de previsão. Em primeiro lugar, é de elevada complexidade conseguir que dados com um elevado grau de importância estratégica, como estes dados referentes ao histórico de vendas de um retalhista, sejam facultados a instituições externas à própria empresa, mesmo que seja para efeitos de pesquisa, uma vez que ao facultar dados deste carácter a empresa pode estar a criar uma vulnerabilidade caso os mesmos extravasem os limites da confidencialidade. Pelo que para que esta dissertação pudesse ser realizada foi necessário seguir um protocolo de confidencialidade celebrado entre ambas as partes. É de acrescentar ainda que este conjunto de dados tem proveniência de um retalhista, conforme já enunciado, facto que distingue este estudo de todos os outros disponíveis até ao presente momento. Analisando toda a literatura existente na área de previsão de produtos de carácter intermitente rapidamente compreendemos que praticamente a totalidade desses trabalhos incide sobre os mesmos conjuntos de dados temporais, obtidos através da *Royal Air Force* (RAF) ou pertencentes a uma empresa de produção e comercialização de partes de peças para aviões. Assim, este é o primeiro trabalho na área que incide na procura de SKUs pertencente a um retalhista.

Em segundo lugar, é recorrentemente complicado obter um conjunto de dados com um volume considerável que permita que os métodos de previsão sejam aplicados, treinados e

testados, de forma a que os resultados de previsão obtidos não sejam comprometidos. Este é um tópico com um relevo tão significativo que foi já tema de investigação e tem uma sustentabilidade teórica bastante extensa. De uma forma sucinta o que acontece é que quando efetuamos estudos de previsão de procura é seguida a estratégia de fraturar o conjunto de dados de que dispomos em duas partes, uma que serve para treinar os modelos de previsão, *fitting* do modelo, e que é conhecida como *in-sample data* e uma segunda parte que serve para medir quão bom é o comportamento de previsão do modelo, *testing* do modelo, e que é conhecido como *out-of-sample data*. Tipicamente, a estratégia de seleção da zona de corte do conjunto de dados é feita numa percentagem de 80% - 20%, respetivamente, para o *in-sample* e o *out-of-sample*. É aqui que a problemática da extensão do conjunto de dados pode ser relevante, uma vez que o tamanho do *out-of-sample* não deverá ser inferior ao máximo horizonte de previsão estabelecido. Assim, estas percentagens poderão ter que ser alteradas em função do tamanho do nosso conjunto de dados e da extensão de previsão que pretendemos realizar, existindo, assim, uma maior flexibilidade de decisão para o analista tanto quanto maior for o conjunto de dados que este possui. A escolha destes valores vai ao encontro do defendido por Hyndman et al. (2013) onde é assumida como recorrente a utilização do valor de 20% da amostra total de dados, para *out-of-sample*, para efeitos de teste dos resultados obtidos. Por último, a estrutura com a qual estes dados são fornecidos também é um fator com bastante importância na análise a efetuar e um dos primeiros passos a ter em conta quando se realiza este tipo de trabalho uma vez que à partida cada empresa terá um formato de armazenamento dos seus dados potencialmente diferente das restantes e, com elevada probabilidade, esse formato em bruto não permite ao analista realizar o seu trabalho e aplicar os métodos que pretende. Pelo que, o passo inicial é estudar os dados que são fornecidos, entendê-los e processá-los para um formato que permite a posterior análise. Esta fase apesar de não conferir nenhum valor acrescentado ao trabalho é muito demorada e a sua realização é de elevado relevo para o sucesso dos passos subsequentes.

Os dados dos quais dispomos foram fornecidos em bruto, no mesmo formato com que são geridos internamente pela JM, tendo sido posteriormente desenvolvido um trabalho de tratamento anterior à realização das nossas previsões por forma a tornar este material bruto num conjunto de dados tratável. Deste tratamento de dados resultou um conjunto de dados organizados por loja, e dentro de cada loja uma separação por categorias de produtos, código de categoria e designação, e dentro de cada categoria uma divisão por SKUs. Foi também realizada uma agregação temporal dos dados passando os valores registados a serem referentes a vendas semanais por SKU ao invés de vendas diárias por SKU. É de notar que a realização deste tipo de processos implica obrigatoriamente uma perda de informação acerca da série em análise, contudo, os benefícios obtidos, diminuição do grau de intermitência dos dados, são significativamente mais importantes para a análise levada a cabo do que a perda obtida ao realizarmos este procedimento. Contudo, não se deve deixar de notar que fatores como a sazonalidade dentro da própria semana ou entre dias homólogos entre semanas são perdidos ao realizarmos esta agregação dos dados para dados semanais.

Na Tabela 10 é apresentada a estrutura que categoriza os dados que foram facultados para a execução do estudo levado a cabo no INESC TEC. Conforme podemos averiguar, o retalhista separa os SKUs de uma forma que é lógica em termos de gestão da sua atividade, sendo que para o efeito estes estão agrupados em famílias de produtos de acordo com as suas características principais e funcionalidades. Nesta tabela são apresentados o número de SKUs

de cada uma das áreas de produtos existentes, bem como a percentagem que esses valores representam no volume total de SKUs existentes nas 412 lojas.

Após uma análise a estes valores, foram selecionadas as 6 áreas mais representativas: perecíveis especializados, mercearia, perecíveis não especializados, produtos pessoais, detergentes e produtos de limpeza e bebidas. Estas áreas totalizam 93% do volume total das vendas diárias nas 412 lojas entre 3 de Janeiro de 2012 e 27 de Abril de 2015, num total de 1211 dias que posteriormente foram agregados em 173 semanas.

Tabela 10 - Número de SKUs por área de atividade.

Número de SKUs por área		
Área	Nº de SKUs	Percentagem
Perecíveis especializados	6302	4,9%
Mercearia	6217	4,8%
Perecíveis não especializados	3682	2,9%
Produtos pessoais	3606	2,8%
Detergentes e produtos de limpeza	2514	1,90%
Bebidas	1995	1,50%
Subtotal	24316	18,8%
Têxtil	84044	65%
Bazar ligeiro	9778	7,6%
Elétricos e entretenimento	6202	4,8%
Negócios complementares	4302	3,3%
Têxtil JM	339	0,26%
Serviços	86	0,0666%
Artigos retornáveis	11	0,0085%
Economato	3	0,002324%
Subtotal	104765	81,2%
TOTAL	129081	100%

Os SKUs existentes na loja que foi alvo do nosso estudo, a de maior dimensão das disponíveis, foram agrupados de acordo com a existência ou não de zeros nas suas séries de vendas semanais. E para este trabalho foram selecionados todos os SKUs que apresentaram uma percentagem de semanas com procura nula dentro do intervalo [50%, 100%]. Esta percentagem corresponde a 4991 SKUs. Na Tabela 11 apresenta-se uma estatística descritiva desse conjunto de SKUs.

É de notar o seguinte facto quanto aos valores não abrangidos, ou seja, os valores que se encontram na gama de procura nula de [0%, 50%], estes representam quase na totalidade SKUs que se inserem em duas situações e cujo estudo não tem relevo para o trabalho aqui desenvolvido. Basicamente, nesta gama de valores, temos presentes SKUs que saíram de linha ou SKUs que entraram em linha. Quanto ao primeiro caso, é simples compreender que o seu estudo não tem qualquer relevo prático para o retalhista uma vez que estes deixam de estar

disponíveis nas suas lojas. Porém o segundo caso tem muito interesse em ser estudado. Contudo, o estudo de previsão de produtos que entraram em linha, pela particularidade da falta de histórico de dados de vendas que possibilite a aplicação de métodos de previsão, implica o recurso a outros métodos e estratégias de previsão que saem do espetro de aplicação do presente trabalho. Fundamentalmente, por forma a ser possível realizar previsões sobre este ultimo tipo de SKUs é necessário recorrer ao histórico de dados de SKUs que apresentem um grau de similaridade tão elevado que os seus dados possam completar a série daquele que está a entrar em linha por forma a extrapolar-se previsões de consumo desse SKU. Assim, estes factos contextualização a escolha pelo intervalo de valores abrangido por esta dissertação.

Tabela 11 - Estatística descritiva do conjunto de dados.

4991 SKUs	Quantidade de procura		Intervalo entre procura		Quantidade de procura por período	
	(unidades)		(semanas)		(unidades/semana)	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
Mínimo	5,7	140,4	1,0	0,0	3,8	139,9
Percentil 25	23,3	220,3	1,0	0,0	17,6	215,9
Mediana	31,2	235,7	1,0	0,0	28,0	233,2
Percentil 75	42,1	254,3	1,0	0,2	39,1	251,8
Máximo	175,2	514,3	19,3	22,5	173,5	513,4

3.3. Procedimento experimental

3.3.1. Processo de rolling

O procedimento experimental que foi realizado para a aplicação dos métodos SES, CRO, SBA, TSB e iMAPA e modelos de espaço de estados CRO, SBA, TSB, Auto e Fixed aos dados já preparados baseia-se num processo de *rolling*, por forma a avaliar a capacidade de previsão de cada um dos métodos. O programa para o efeito foi desenvolvido na linguagem de programação R tirando partido dos pacotes “*tsintermittent*” e “*smooth*”.

Com esta análise perspetiva-se a possibilidade de estabelecer algumas relações entre os SKUs em análise e os métodos que melhor comportamento apresentam na previsão da sua procura com a eventual possibilidade de os agrupar em *clusters* tendo em vista dotar o retalhista de *know-how* que lhe permita a aplicação dos métodos aos *clusters* estabelecidos viabilizando o processo de previsão através da drástica redução do tempo de análise e consumo de recursos computacionais para o efeito. É de fazer notar o peso temporal que acarreta levar a cabo um procedimento deste carater e a quantidade de recursos computacionais necessários para que essa análise possa ocorrer num período de tempo que seja considerado aceitável. Inicialmente os processamentos foram realizados por recurso a computadores pessoais com a dedicação de 6 *cores*. Contudo, muito rapidamente

compreendemos que esta solução não seria viável nem suficiente para realizar um processamento destas dimensões. A solução encontrada para este problema foi realizar o processamento na infraestrutura de nós disponível no serviço GridFEUP.

Por forma a obter uma previsão, ou conjunto de previsões de procura, partindo do histórico de um conjunto de dados compreendidos entre $t = 1$ e um tamanho especificado $t = n$, é possível realizar dois tipos de processos de previsão. Podemos efetuar previsões a partir de uma origem comum nos dados, sendo este processo denominado de previsão de origem fixa, ou podemos realizar uma atualização do ponto a partir do qual as previsões são realizadas mantendo um horizonte temporal de previsão constante, denominando-se este último processo de previsão de origem variável, ou processo de *rolling*.

Dos dois processos em análise aquele que se apresenta de maior interesse para os analistas é o processo de *rolling*, uma vez que este processo permite obter um maior número de erros de previsão.

No caso da utilização do processo de *rolling*, existe ainda uma decisão a ser tomada antes da sua utilização, que é a seleção do horizonte temporal de previsão. Usualmente, a escolha divide-se entre realizar previsões com um horizonte temporal de uma unidade, $h = 1$, ou optar por uma previsão com um horizonte temporal mais longínquo, $h > 1$.

O procedimento de previsão de um processo de *rolling*, com horizonte de previsão igual a um instante, pode ser visto na Figura 13, e é formalmente descrito como um processo de três etapas. A primeira compreende a determinação do conjunto de treino que deverá compreender as observações que vão do instante 1 até ao período imediatamente anterior a ser previsto, ou seja $\{1, \dots, k - 1\}$, realiza-se a previsão para o instante k e determina-se o erro associado à previsão. A segunda etapa é a integração do período já previsto, ou seja, passamos a ter $k = k + 1$, e repete-se o ponto anterior até atingir a previsão do último valor pretendido. A última etapa corresponde à determinação das medidas de erro de previsão obtidas a partir dos erros calculados nos vários passos do processo. Denomina-se processo de previsão de origem variável, uma vez que a origem a partir da qual é feita a previsão varia ao longo do processo.

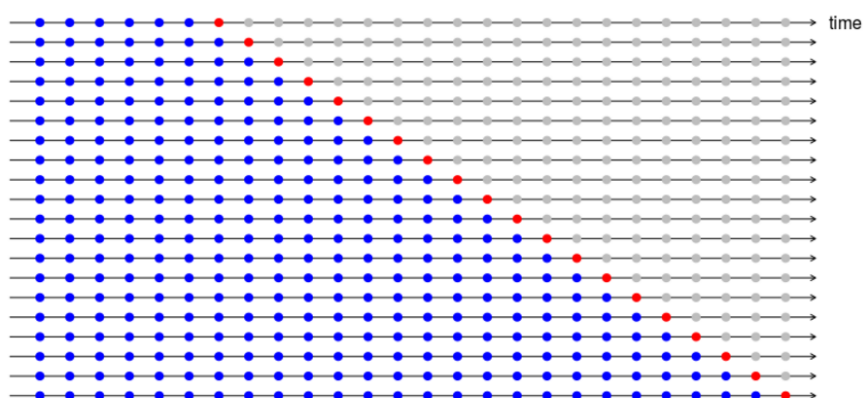


Figura 13 - Processo de rolling com previsão a 1-passo.

O procedimento de previsão com um processo de *rolling* com horizonte superior a uma unidade temporal é bastante similar ao descrito para o caso de $h = 1$, com a única diferença de que a previsão que é realizada não é para o valor seguinte, mas para um valor h períodos à frente. Ou seja, seguindo o exemplo da Figura 14, na primeira linha é feita a previsão para a

posição $k + h$, sendo neste caso $h = 4$. É realizado o treino do modelo, a previsão, a determinação do erro, a integração do período subsequente, repetindo-se o processo descrito para todos o conjunto de dados e no final são determinadas as medidas de precisão de previsão em função dos erros calculados.

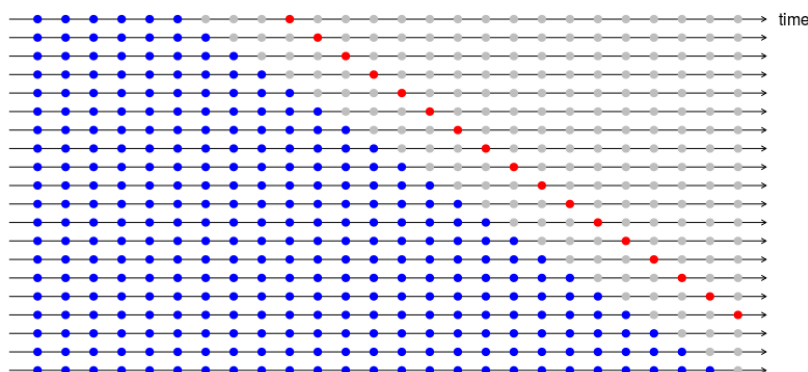


Figura 14 - Processo de rolling com horizonte de previsão de quatro períodos.

Decidimos por esta abordagem de previsão com um horizonte temporal de um instante, com realimentação e re-estimação do método em cada passo, tentando simular o procedimento rotineiro do retalhista e dessa forma avaliar qual o método que melhor se adequa à sua atividade. Foi considerado um conjunto de teste de 52 semanas tendo em vista essa avaliação para o período de 1 ano.

Após serem realizadas todas as iterações do *rolling*, no nosso caso 52 iterações correspondentes a 52 previsões, para cada método, foram calculadas diversas medidas de erro de modo a que estes pudessem ser no final comparados em termos de eficiência de previsão.

3.3.2. Métodos e pacotes de software

O pacote de software “*tsintermittent*” para previsão de séries temporais intermitentes existente no *software* livre R foi desenvolvido por Nikolaos Kourentzes e Fotios Petropoulos em 2016. Segundo estes, o pacote é constituído por funções para análise e previsão de procura intermitente, ou seja, para séries temporais de *slow-moving items*.

Dos modelos existentes neste pacote e que foram utilizados na nossa análise de dados consta o Croston, o SBA, o TSB, o SES e o iMAPA.

Para a implementação destes métodos Kourentzes e Petropoulos basearam-se na especificação dos modelos e conceitos originalmente propostos pelos respetivos autores e que já foram apresentadas no capítulo 2. A título de exemplo, a função Croston existente neste pacote invoca a implementação original em R deste método.

Este pacote é bastante interessante na perspetiva de que realiza a otimização dos valores iniciais e parâmetros de alisamento de cada um dos métodos. A implementação realizada para essa otimização é, segundos os autores do pacote, a apresentada no trabalho Kourentzes (2014). Fundamentalmente, como já foi brevemente referido no capítulo 2, o que é apresentado nesse trabalho como metodologia de otimização de valores iniciais é o recurso à minimização das funções custo MSE e MAE, sendo ainda propostas pelo próprio Kourentzes

duas novas funções custo, que demonstram melhor desempenho para realizar a otimização destes métodos quando comparadas com o desempenho que se consegue obter da otimização do MSE e do MAE. As funções que ele propõe são o MSR e o MAR. Em termos de implementação é dada a possibilidade ao utilizador de optar por uma otimização realizada por qualquer uma das quatro funções custo apresentadas, sendo que por defeito o programa utiliza o MAR.

Porém, penso que é relevante referir que a implementação realizada pelo iMAPA para dados intermitentes realiza uma seleção automática do modelo a aplicar (SES, CRO ou SBA) ao conjunto de dados por recuso a uma classificação PK (Petropoulos Kourentzes). A exposição deste modelo de classificação foi já anteriormente introduzida no capítulo anterior na secção que aborda o iMAPA, mais especificamente quando se analisa as metodologias de seleção de métodos e a forma de os classificar.

Este pacote continua em constante atualização

O pacote *Smooth* implementa em R os modelos de espaço de estados e foi desenvolvido por Svetunkov (2017).

Especificamente para dados intermitentes, implementa os métodos de Croston, SBA, TSB, Auto e Fixed. Ao embeber estes métodos na *framework* dos modelos de espaço de estados tira partido das vantagens que a mesma proporciona nomeadamente a otimização dos parâmetros e valores iniciais através da maximização da função de verosimilhança (função custo), que se espera proporcionar melhores resultados de previsão.

O pacote *smooth* apresenta um nível de utilização mais amplo do que o pacote *tsintermittent* não só pela função custo, mas também por proporcionar a utilização simultânea dos modelos de espaço de estados para *fast-moving goods*.

O *smooth* é um pacote que apenas foi introduzido como ferramenta de previsão este ano, não existindo atualmente disponível literatura que apresente resultados comparativos entre a aplicação de ambos os pacotes a série temporais, tornando, este facto, o presente trabalho de dissertação o primeiro a ser realizado com este foco de estudo.

3.3.3. Medidas de avaliação de previsão

Após a execução de uma previsão é habitual realizar-se uma avaliação da performance do método na geração dessa previsão e isso é determinado por recurso ao cálculo do erro que o método cometeu ao gerar esse valor. Para o efeito são utilizadas várias medidas de avaliação desses erros.

Quando analisamos dados intermitentes, que relembro, são conjuntos de dados onde se registam valores nulos, é necessário efetuar uma análise cuidadosa na escolha dessas métricas uma vez que nem todas são adequadas para dados intermitentes podendo ocorrer, por exemplo, situações de indeterminação que inviabiliza a utilização dessas medidas. Em suma, da mesma forma que é necessário recorrer a métodos de previsão específicos para dados com intermitência, o mesmo se aplica quando selecionamos medidas de avaliação das previsões.

Assim, para a análise dos erros associados às previsões que obtivemos, recorreremos a 10 medidas de erro diferentes com o objetivo de que no seu conjunto estas afirmem as características do conjunto de erros.

Se analisarmos rapidamente as medidas de erro que selecionamos para serem aplicadas no nosso caso de estudo, todas elas são medidas de erro ditos de escala. No fundo o erro

escalado é um erro relativo cujo objetivo é a remoção da escala do erro para que os erros das diferentes séries temporais que compõem o conjunto de dados possam ser comparados.

Passando à análise das medidas de erro especificamente utilizadas neste trabalho, em primeiro lugar utilizamos o *Mean scaled Error* (MsE) que é uma medida permite aferir o enviesamento dos erros:

$$MsE = mean \left(\frac{y_{N+h} - \hat{y}_h}{\frac{1}{N} \sum_{t=1}^N y_t} \right) \quad (68)$$

Onde y_{N+h} corresponde ao valor observado no instante $t+h$, fazendo-se uma previsão para o instante h . Fundamentalmente, em cada passo do *rolling*, é calculado o erro da previsão, que resulta da diferença entre o valor efetivamente observado de procura e o valor previsto, é calculada a média do conjunto de treino em causa, e por fim é feita a divisão do primeiro pelo segundo por forma a obter o erro escalado pela média do conjunto de treino. Posteriormente, é feita a média dos 52 erros obtidos com a realização de todos os passos do processo de *rolling*. Estas designações mantêm-se válidas para as restantes fórmulas apresentadas em seguida.

Se ao invés da média determinarmos a mediana dos 52 erros referidos obtemos o *Median scaled Error* (MdsE):

$$MdsE = median \left(\frac{y_{N+h} - \hat{y}_h}{\frac{1}{N} \sum_{t=1}^N y_t} \right) \quad (69)$$

Em terceiro lugar, utilizamos o *Mean scaled Absolute Error* (MsAE) que resulta da média do sAE (*scaled Absolute Error*) escalado sendo dado por:

$$MsAE = mean \left(\frac{|y_{N+h} - \hat{y}_h|}{\frac{1}{N} \sum_{t=1}^N y_t} \right) \quad (70)$$

Esta medida de erro permite-nos aferir a exatidão dos erros obtidos em cada um dos métodos. O seu cálculo é em tudo similar ao MsE com a diferença de que no cálculo do erro este ser cálculo ser realizado para o valor absoluto, ou seja, pelo cálculo do módulo da diferença entre o valor efetivamente observado de procura e o valor previsto. Se ao invés da média determinarmos a mediana dos erros referidos obtemos o *Median scaled Absolute Error* (MdsAE):

$$MdsAE = median \left(\frac{|y_{N+h} - \hat{y}_h|}{\frac{1}{N} \sum_{t=1}^N y_t} \right) \quad (71)$$

Quando comparada com a mediana, a média permite-nos aferir a existência de valores extremos de erros.

Em quinto lugar, utilizamos o *Mean scaled Squared Error* (MsSE) que resulta da média dos *scaled Squared Error* (sSEs) calculados ao longo do *rolling* e que permite avaliar a variância dos erros obtidos, sendo calculado por:

$$MsSE = mean\left(\left(\frac{y_{N+h} - \hat{y}_h}{\frac{1}{N} \sum_{t=1}^N y_t}\right)^2\right) \quad (72)$$

Se ao invés de a média determinarmos a mediana dos erros referidos obtemos o *Median scaled Squared Error* (MdsSE):

$$MdsSE = median\left(\left(\frac{y_{N+h} - \hat{y}_h}{\frac{1}{N} \sum_{t=1}^N y_t}\right)^2\right) \quad (73)$$

Em sétimo lugar, recorreremos ao *Mean Absolute Scaled Error* (MASE) resultante do cálculo da média do ASE (Absolute Scaled Error) escalado com o erro absoluto médio das previsões do método *Naïve* para o conjunto de treino, ou seja:

$$MASE = mean\left(\left|\frac{y_{N+h} - \hat{y}_h}{\frac{1}{N-1} \sum_{t=2}^N |y_t - y_{t-1}|}\right|\right) \quad (74)$$

Se ao invés da média determinarmos a mediana dos 52 erros referidos obtemos o *Median Absolute Scaled Error* (MdASE):

$$MdASE = median\left(\left|\frac{y_{N+h} - \hat{y}_h}{\frac{1}{N-1} \sum_{t=2}^N |y_t - y_{t-1}|}\right|\right) \quad (75)$$

A realização do correto procedimento de avaliação da performance dos métodos no ato de previsão é de extrema importância uma vez que é a forma de o analista ter uma noção de quão corretas e quão próximas da realidade estão as previsões obtidas, tendo sempre em consciência que um modelo que tem uma boa adaptação aos dados, sendo denominado processo de *fitting* o processo de treino do modelo, não implica obrigatoriamente que venha a gerar boas previsões para o futuro.

Por último, e numa fase de conclusão das experimentações computacionais realizadas, foram ainda introduzidos nas metodologias de avaliação de erro o MsPIS e o MsAPIS.

Estes erros são interessantes introduzir na avaliação dos métodos uma vez que se baseiam no PIS, erro que apresenta a possibilidade de obter uma interpretação prática para o retalhista das previsões que se obtêm.

Este erro está inserido na categoria de medidas relacionadas com o inventário, sendo um método específico para situações de previsão relacionadas com procura, que é o caso do nosso caso de estudo, e avalia medidas de erro de previsão associadas com performance de inventário. O método é conhecido por *Periods in stock* (PIS), é apresentado na Equação (76) e este baseia-se na assunção de um *stock* fictício ao qual é entregue os valores previstos em cada período de tempo das nossas previsões de cada método e, nesse período temporal específico, as unidades definidas são removidas do stock.

$$PIS = -\sum_{i=1}^n \sum_{j=1}^i (D_j - F_j) \quad (76)$$

A Tabela 11 apresenta um exemplo prático da aplicação do PIS. Imaginando que a nossa curta série temporal é constituída por um conjunto de 10 dados, com o período a ser apresentado na primeira coluna, seguido da determinação do PIS de cada período, a previsão realizada para esse período e por fim a procura que realmente se verifica, ou seja, a nossa série temporal de dados de procura por esse SKU.

Tabela 12 - Exemplo de aplicação PIS

Períodos	PIS	Previsão	Procura
0			
1	2	2	0
2	3	1	0
3	8	3	3
4	14	4	1
5	22	2	0
6	25	0	5
7	30	5	3
8	36	1	0
9	42	2	2
10	-7	1	56

Para obtermos alguma sensibilidade de como se processa o cálculo do PIS é demonstrado na Equação (77) como se obteve, por exemplo, o PIS_5 .

$$\begin{aligned}
 PIS_5 &= -\sum_{i=1}^5 \sum_{j=1}^i (D_j - F_j) = \\
 &= -[(0 - 2) + ((0 - 2) + (0 - 1)) + \\
 &\quad ((0 - 2) + (0 - 1) + (3 - 3)) + ((0 - 2) + (0 - 1) + (3 - 3) + \\
 &\quad (1 - 4)) + \\
 &\quad ((0 - 2) + (0 - 1) + (3 - 3) + (1 - 4) + (0 - 2))] = 22
 \end{aligned} \tag{77}$$

O valor obtido do PIS calculado para um determinado período é interpretado da seguinte forma, se o valor calculado for superior a 0, então isto significa que houve uma sobre de skus em stock, por outro lado se o valor calculado for inferior a 0 implica que aconteceu o fenómeno de stock-out.

Neste trabalho foram aplicadas duas versões diferentes do PIS. Utilizou-se o MsPIS (*scaled Mean Periods in Stock*), que é o erro escalado médio do Períodos em *Stock*, e o MsAPIS, que é o erro escalado absoluto médio do Períodos em *Stock*.

O recurso ao MsPIS permite-nos recorrer a uma medida média que está escalada pelo que nos possibilita realizar uma análise de uma medida de erro que é independente da escala em que estamos a trabalhar. Fundamentalmente, o MsPIS corresponde à media simples calculada ao longo de todas as séries do PIS escalado pelo valor médio do *in-sample*.

$$MsPIS = mean \left(\frac{-\sum_{i=1}^n \sum_{j=1}^i (D_j - F_j)}{\frac{1}{N} \sum_{t=1}^N y_t} \right) \tag{78}$$

Foi ainda alvo de utilização a medida de erro MsAPIS (*scaled Mean Absolue Periods in Stock*), que é o erro absoluto escalado médio dos períodos em *stock*.

Fundamentalmente a diferença, em termos matemáticos, entre o MsPIS e o MsAPIS é a determinação do erro absoluto em vez de apenas o erro escalado.

$$MsAPIS = mean\left(\frac{\left|-\sum_{i=1}^n \sum_{j=1}^i (D_j - F_j)\right|}{\frac{1}{N} \sum_{t=1}^N y_t}\right) \quad (79)$$

3.4. Resultados

Na tabela 13 e nos anexos A6, A7 e A8 estão apresentados os resultados obtidos para cada um dos métodos definidos, sendo referenciados os valores obtidos para cada um dos erros que foram tidos em conta e acompanhando os resultados com um *ranking* que foi desenvolvido com o intuito de seriar os resultados.

Analisando a tabela 13, que corresponde ao *ranking* global, podemos visualizar nesta tabela que os modelos de espaço de estados para séries intermitentes disponibilizados no pacote *smooth* tiveram todos prestação superior aos métodos homólogos que são implementados no pacote *tsintermittent*.

Tabela 13 - *Rank* médio de cada método para os diferentes horizontes e *Rank* global

Método	Horizonte temporal						Média		Rank Global
	1	2	3	4	6	8	1-4	1-8	
TSBs	3.200	3.600	3.600	3.600	3.400	3.300	3.500	3.450	3.450
NAIVE	3.600	3.600	3.800	3.800	3.900	4.100	3.700	3.825	3.825
AUTOs	4.100	4.300	4.300	4.300	4.100	4.100	4.250	4.188	4.188
iMAPA	5.600	5.400	5.500	5.500	5.700	5.600	5.500	5.556	5.556
TSBi	6.000	5.950	5.700	5.800	5.500	5.650	5.863	5.744	5.744
SBA _s	6.300	6.400	6.700	6.700	6.800	6.800	6.525	6.663	6.663
SESi	7.300	7.100	6.950	6.700	6.700	6.750	7.013	6.863	6.863
SBAi	6.800	7.000	7.050	6.950	7.250	7.200	6.950	7.025	7.025
FIXEDs	7.050	7.000	6.700	6.850	7.300	6.900	6.900	7.063	7.063
CROSTONs	7.600	7.450	7.500	7.500	7.600	7.600	7.513	7.556	7.556
CROSTONi	8.450	8.200	8.200	8.300	7.750	7.400	8.288	7.994	7.994

Assim, podemos verificar que no topo, e implementados pelo espaço de estados temos o TSB e o Auto e na base da tabela estão os Croston dos dois pacotes utilizados.

O método *Naïve* surge numa posição bastante interessante no ranking da Tabela 13. Conforme podemos verificar ele surge em segundo lugar na posição global. Demonstrando melhor prestação que todos os métodos do pacote *tsintermittent* e ainda melhor desempenho que o Croston e o SBA implementados pelo *Smooth*. Este desempenho tão elevado de um método tão simples pode ser explicado pelo grau de intermitência dos conjuntos de dados, uma vez que tendo dados com elevado número de zeros que surgem recorrentemente de forma sequencial nas observações, este método para esses casos vai ter um erro zero na realização das suas previsões, o que automaticamente lhe atribui vantagem face aos restantes e o catapulta para as posições superiores do *ranking*.

O *ranking* registado nesta tabela trata-se de um *ranking* cujos valores são obtidos pela determinação da média dos valores de posicionamento dos modelos em cada erro, ou seja, é realizado uma seriação da posição por ordem de *performance* de 1 a 11 por erro, efetuado um somatório das posições obtidas por modelo e o total é dividido pelo número de erros.

Podemos também verificar que o TSB de entre os diversos métodos é aquele que demonstra melhor prestação geral, este facto poderá estar associado com a característica que o próprio modelo possui de atualizar, continuamente ao longo dos diversos períodos, o valor da estimação de probabilidade de procura.

Em relação à função Auto que é disponibilizada pelo pacote *Smooth*, e que executa uma seleção automática do modelo a aplicar por recurso ao cálculo do AIC, que seria de esperar ser o candidato mais forte ao primeiro lugar da performance entre os diversos métodos, este fica colocado em segundo, excluindo desta reflexão o *Naïve* cuja posição já foi contextualizada, sendo ultrapassado em desempenho pelo TSB do *Smooth*. Este resultado reforça o facto conhecido de que um modelo que se ajusta bem aos dados não implica que seja o modelo que possua o melhor desempenho em termos de previsão.

Em ambos os pacotes, e como seria esperado, o Croston tem um pior desempenho quando comparado com o método que foi gerado para a correção do seu enviesamento, ou seja, o SBA, sendo que o CROs obteve melhor performance que o CROi.

O posicionamento do iMAPA vai ao encontro do que se poderia esperar deste método, uma vez que este é baseado na aplicação do SBA, CRO e do SES e recorre à seleção do método mais adequado a aplicar em função do conjunto de dados, apresentando melhor desempenho que os 3 métodos individualmente.

Na base da nossa tabela podemos verificar que estão posicionados os Croston sendo que o SES surge em sétimo lugar, quando seria de esperar que o Croston tivesse uma performance superior à do SES. Contudo, com a otimização de valores e parâmetros iniciais que o *tsintermittent* realiza antes de aplicar os modelos, o SES demonstra ter uma performance superior na previsão. Este facto ainda não tinha sido registado nos trabalhos de sustentação destes métodos realizados até ao presente momento uma vez que nestes não era realizada otimização de parâmetros.

Verifica-se ainda que o modelo Fixed disponibilizado pelo *Smooth* apresenta um dos piores desempenhos, estando inclusive ao nível do desempenho obtido pelo Croston, sendo superado por todos os métodos do *tsintermittent* à exceção do Crostoni.

Pode ainda acrescentar-se que as medidas de erro apresentados nesta tabela são razoavelmente consistentes entre si.

Capítulo 4

Conclusões e Trabalho Futuro

Conforme foi possível constatar pela análise da literatura existente, da orientação que seguem os trabalhos desenvolvidos ao longo das últimas décadas e o sentido que tomam aqueles trabalhos que ganham destaque na atualidade na área de previsão, a previsão de procura baseada em dados com intermitência e o desenvolvimento e aplicação de métodos que lhe são específicos está a ganhar cada vez maior dimensão e a prender o foco dos investigadores desta área, apesar do estudo da intermitência mostrar visivelmente margem para desenvolvimento.

Conjuntos de dados como aquele de que foi objeto este trabalho, que apresentam elevadíssimos graus de intermitência, provam a urgência que existe em continuar a aprofundar cada vez mais este assunto no sentido de dotar o mercado de ferramentas estratégicas que permitam previsões rápidas e eficientes sem comprometerem a exatidão dos resultados obtidos e que consigam transmitir confiança aos utilizadores para se poderem basear nestes resultados e com isto especificarem as operações das suas empresas de modo a acompanharem a procura e o mercado.

Faz-se também notar que os resultados obtidos, bem como as conclusões daí decorrentes, são referentes à análise de apenas uma das lojas Pingo Doce do grupo Jerónimo Martins, pelo que a extrapolação para as restantes lojas do grupo poderá ser abusiva e levar a erros. Apesar da loja selecionada ser a que apresenta maior dimensão e maior volume de vendas do Pingo Doce, este facto não implica que as conclusões retiradas da sua análise tenham que ser obrigatoriamente extensíveis às restantes lojas.

Um dos tópicos de incidência deste trabalho foi uma análise comparativa da eficiência de previsão dos métodos SES, CRO, SBA, TSB, iMAPA e modelos de espaço de estados próprios para previsão de séries temporais com elevado grau de intermitência.

Esta análise comparativa foi levada a cabo através da previsão de um grande número de séries temporais intermitentes relativas a vendas no setor do retalho da distribuição alimentar, com recurso a um processo de *rolling* que aumentou o número de erros disponíveis e aferiu a capacidade de avaliação desses métodos num período temporal de 1 ano, tentando simular a atividade rotineira do retalhista. A implementação deste procedimento experimental tirou partido da implementação destes métodos existente nos pacotes *tsintermittent* e *smooth* do *software* estatístico R.

Dos resultados obtidos pode concluir-se que o recurso aos modelos de espaço de estados permite obter uma clara superioridade em termos de *performance* face à implementação dos métodos intermitentes homólogos, devendo-se sobretudo à estimação de parâmetros que cada um dos pacotes realiza. Esta superior eficiência na estimação de parâmetros deve-se ao facto de o *Smooth* nesse processo recorrer à maximização da função de verosimilhança, ou minimização do MSE que ele prova serem similares, enquanto que o *tsintermittent* realiza a minimização do MAR e que, pela análise dos resultados obtidos se tornou claramente determinante para a performance do método. Em suma, o recurso ao modelo de espaço de estados para implementação, conforme realiza o *Smooth*, apresenta uma *performance* superior uma vez que é uma *framework* que permite realizar uma estimação de parâmetros e seleção dos modelos mais adequados a serem aplicados, permite realizar uma comparação da adequação dos modelos através da determinação do AIC e possibilita ainda a determinação de intervalos de confiança uma vez que é uma *framework* de erro aleatório.

Concluiu-se ainda que o modelo FIXED implementado pelo modelo de espaço de estados não é uma boa metodologia de previsão a seguir, demonstrando que a previsão com recurso a probabilidades fixas de ocorrência não é a melhor opção.

Verifica-se, conforme seria de esperar, que um método tão simples como o *Naïve* apresenta desempenhos interessantíssimos no caso de séries de dados com elevado grau de intermitência nos dados, sendo que se verifica que o *benchmark* ultrapassa o desempenho de outros métodos mais desenvolvidos e mais complexos.

A necessidade de um método como o Auto que o *smooth* implementa é clara, ou seja, a possibilidade de ter um método que realize uma seleção automática dos modelos a aplicar em função do conjunto de dados em análise e tendo em conta o grau de intermitência dos mesmos. Uma ferramenta como este método permitiria aos analistas ter mais flexibilidade dos métodos aplicados ao conjunto de dados, retirando o melhor da performance obtida da correta seleção do método ao conjunto de dados em questão.

Originalmente, quando este trabalho de dissertação foi iniciado, para além de se perspetivar realizar a comparação entre implementação dos métodos pela definição (*tsintermittent*) vs implementação por recurso ao espaço de estados (*Smooth*) foi também idealizado desenvolver uma análise da previsão de procura com este conjunto de dados com integração de regressores.

Esta metodologia de análise referida permite realizar as previsões dos múltiplos métodos, neste caso existentes no pacote *Smooth*, e já abordados, fornecendo-lhes informação fulcral à previsão acerca das séries temporais em análise. Informação essa previamente selecionada e considerada de relevo para o processo de previsão. Dos métodos implementados no *smooth* e que seriam testados com recurso a regressores conta-se o *Crostonx*, *SBx*, *TSBx*, *Autox* e *Fixedx*. Após uma cuidada análise acerca dos fatores que influenciam a procura dos sku's deste retalhista, e que podem ser extrapolados por similaridade a outros retalhistas ou até mesmo outras áreas de investimento, chegou-se à conclusão da existência de 50 regressores que têm que ser tidos em conta para a realização desta simulação, sendo eles o preço, desconto relativo, dias de promoção numa semana, primeira e ultima semana do mês e eventos de calendário.

O desenvolvimento destas simulações tem um peso de programação muito significativo, recorre a um pacote extremamente recente e que neste campo da regressão não está ainda totalmente limpo de *bugs* de programação e envolve estratégias específicas (ex: PCA) para redução da quantidade de parâmetros fornecidos ao método uma vez que para 50 regressores

as simulações invariavelmente chegam a crash pela enorme quantidade de parâmetros fornecidos. É de referir ainda que estamos a falar de simulações extremamente pesadas em termos de processamento computacional, com necessidade de recursos de processamento muito complicados de obter mesmo para a GridFEUP, o que dificultou ainda mais a possibilidade de execução deste trabalho em tempo útil para ser integrado nesta tese de mestrado.

Porém, é de sublinhar o quão imperativo é realizar uma análise destas a um conjunto de dados como aquele ao qual tivemos acesso. A análise de previsão com recurso a regressores vai dotar os retalhistas de previsões mais sensíveis à sua realidade, integrando nas previsões conhecimentos relevantes das suas operações e aproximando todo o trabalho teórico-prático realizado na área da previsão de procura intermitente da realidade que os retalhistas vivem nas suas empresas no dia-a-dia, e que é, em suma, o fator mais relevante para eles a retirar de um trabalho como este.

Assim, demonstra-se imperativo dar continuidade, num futuro bastante próximo, ao trabalho aqui apresentado no sentido desenvolver competências de previsão na área de intermitência com inclusão de regressores e possibilitar a aproximação entre o trabalho académico que é desenvolvido na área de previsão e os interesses práticos do mercado.

Referências

- Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: A decomposition approach to forecasting. *International Journal Forecast* 16: 521-530.
- Babai, M.Z., Syntetos, A.A., Teunter, R., 2011. Intermittent Demand Estimators: Empirical Performance and Sensitivity to the Smoothing Constants Used. Working paper no. 138-11, Centre de recherche de BEM.
- Babai, M.Z., Ali, M.M., Nikolopoulos, K., 2012. Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega* 40(6): 713-721.
- Banker, S. Demand Forecasting: Going beyond historical shipment data. *Forbes Logistics & transportation*. Disponível em <https://www.forbes.com/sites/stevebanker/2013/09/16/demand-forecasting-going-beyond-historical-shipment-data/#2f49d57f16fb> Acesso em 25/Maio/2017.
- Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. *Operational Research Society* 20(4): 451-468.
- Brown, R.G., 1956. *Exponential Smoothing for Predicting Demand*. Cambridge, Massachusetts. Arthur D. Little Inc., p.15.
- Business Dictionary, Disponível em <http://www.businessdictionary.com/definition/forecasting-system.html> Acesso em 10/Maio/2017.
- Cholette, P.A., 1982. Prior information and ARIMA forecasting. *Journal of Forecasting* 1: 375-384.
- Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5(4): 559-583.
- Corsten D., Gruen T., Stock-outs Cause Walkouts, *Harvard Business Review*. Disponível em <https://hbr.org/2004/05/stock-outs-cause-walkouts> Acesso em 18/Maio/2017.
- Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* (1970-1977), 23(3):289-303.
- Eaves, A.H.C., Kingsman, B.G., 2004. Forecasting for the ordering and stock-holding of spare parts. *The journal of the Operational Research Society* 55(4): 431-437.

- He, C., Xu, X., 2005. Combination of forecasts using self-organizing algorithms. *Journal of forecasting* 24: 269-278.
- Heinecke, G., Syntetos, A.A., Wang, W., 2013. Forecasting - based SKU classification. *International Journal of Production Economics* 143(2): 455-462.
- Hibon, M., Evgeniou, T., 2005. To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting* 21(1): 15-24.
- Hyndman, Rob J., George Athanasopoulos, *Forecasting: principles and practice*, 2013. Disponível em: <https://www.otexts.org/fpp> Acesso em 15/Março/2017.
- Hyndman, Rob J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. *Forecasting with Exponential Smoothing The State Space Approach*, XII, 360 p.
- Johnston, F.R., Boylan, J.E., Shale, E.A., 2003. An examination of the size of orders from customers, their characterization and the implications for inventory control of slow moving items. *Journal Operational Research Society*, 54:833-837.
- Jose, V.R.R., Winkler, R.I., 2008. Simple robust averages of forecasts: some empirical results. *International Journal of Forecasting* 24: 163-169.
- Kolassa, S., 2011. Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting* 27(2): 238-251.
- Kourentzes, N., Petropoulos, F., Trapero, J.R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 30 (2), 29:1-302.
- Kourentzes, N., 2014. On intermittent demand model optimisation and selection. *International Journal Production Economics*, 156:180-190.
- Kourentzes, N., 2013. Intermittent demand forecast with neural networks. *International Journal Production Economics*, 143:198-206.
- Kostenko, A.V., Hyndman, R.J., 2006. A note on the categorization of demand patterns. *Journal of Operational Research Society*. 57, 1256-1257.
- Makridakis, S., Winkler, R., 1983. Average of forecasts: Some empirical results. *Management Science* 29(9): 987-996.
- Nikolopoulos K., Syntetos A.A., Boylan J.E., Petropoulos F. e Assimakopoulos V. (2011). Na aggregate-Disaggregate intermittent demand approach (ADIDA) to forecasting: Na empirical proposition and analysis. *Journal of the Operational Research Society* 62(3): 544-554.
- Petropoulos, F., Kourentzes N., Nikolopoulos, K., 2016. Another look at estimators for intermittent demand. *International Journal Production Economics*, 181:154-161.
- Petropoulos, F., Nikolopoulos, K., Spithourakis, G., Assimakopoulos, V., 2013. Empirical heuristics for improving intermittent demand forecasting. *Ind. Manag. Data Syst.*, 113 (5), 683-696.

- Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. 'Horses for courses' in demand forecasting. *European Journal of Operational Research* 237(1): 152-163.
- Petropoulos, F., Kourentzes, N., 2015. Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66: 914-924.
- Rao, A., 1973. A comment on: forecasting and stock control for intermittent demands. *Operational Research Q.* 24:639-640.
- Romeijnders, W., Teunter, R., van Jaarsveld, W., 2012. A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research* 220(2): 386-393.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., Ducq, Y., 2013. Demand forecasting by temporal aggregation. *Naval Research Logistics* 60(6): 479-498.
- Shale, E.A., Boylan, J.E., Johnston, F.R., 2006. Forecasting for intermittent demand: the estimation of an unbiased average. *Journal Operational Research Society* 57, 588-592.
- Shenstone, L., Hyndman, R.J., 2005. Stochastic Models Underlying Croston's Method for Intermittent Demand Forecasting. *Journal of Forecasting* 24:389-402.
- Snyder, R., 2002. Forecasting sales of slow and fast-moving inventories. *European Journal of Operational Research* 140(3): 684-699.
- Spithourakis, G.P., Petropoulos, F., Babai, M.Z., Nikolopoulos, K., Assimakopoulos, V., 2011. Improving the performance of popular supply chain forecasting techniques. *Supply Chain Forum: An International Journal* 12(4): 16-25.
- Spithourakis, G.P., Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2014. A systemic view of the adida framework. *IMA Management Mathematics* 25(2): 125-137.
- Svetunkov, I., 2017. Statistical models underlying functions of "smooth" package for R, Lancaster Centre for Forecasting.
- Syntetos, A.A., Boylan, J.E., 2001. On the bias of intermittent demand estimates. *International Journal of production Economics* 71(1-3): 457-466.
- Syntetos, A.A., Nikolopoulos, K., Boylan, J.E., Petropoulos, F. and Assimakopoulos, V., 2011. An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. *Journal of the Operational Research Society* 62:544-554.
- Syntetos A.A. and Boylan J.E., 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2):303-314.
- Syntetos A.A. and Boylan J.E., 2006. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics* 103(1): 36-47.
- Teunter, R.H., Duncan, L., 2009. Forecasting intermittent demand: A comparative study. *The Journal of the Operational Research Society* 60(3): 321-329.

- Teunter, R.H., Sani, B., 2009b. On the bias of Croston's forecasting method. *European Journal of Operational Research*, 194(1):177-183.
- Teunter, R.H., Syntetos A.A., Babai, M.Z., 2010. Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research* 203(3): 619-624.
- Teunter, R.H., Syntetos A.A., Babai, M.Z., 2011. Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, 214(3):606-615.
- Timmermann, A., 2006. Forecast combinations. In G.Elliott, C.Granger, & A.Timmermann (Eds.), *Handbook of economic forecasting*, vol. 1 (pg 135-196). Elsevier.
- Wallstrom O., Segerstedt A., 2010. Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics* 128(2):625-636.
- Willemain, T.R., Smart, C.N., Shockor, J.H., DeSautels, P.A., 1994. Forecasting intermittent demand in manufacturing. A comparative evaluation of Croston's method. *International Journal of Forecasting* 10(4): 529-538.

Anexos

A.1 - Combinações da sazonalidade e da tendência

	Seasonal Component		
Trend Component	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A _d (Additive damped)	(A _d ,N)	(A _d ,A)	(A _d ,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
M _d (Multiplicative damped)	(M _d ,N)	(M _d ,A)	(M _d ,M)

A.2 - Classificação proposta por Pegels(1969)

(N,N)	=	simple exponential smoothing
(A,N)	=	Holts linear method
(M,N)	=	Exponential trend method
(A _d ,N)	=	additive damped trend method
(M _d ,N)	=	multiplicative damped trend method
(A,A)	=	additive Holt-Winters method
(A,M)	=	multiplicative Holt-Winters method
(A _d ,M)	=	Holt-Winters damped method

A.3 - Métodos e valores iniciais

Method	Initial values
(N,N)	$\ell_0 = y_1$
(A,N) (A _d ,N)	$\ell_0 = y_1, b_0 = y_2 - y_1$
(M,N) (M _d ,N)	$\ell_0 = y_1, b_0 = y_2/y_1$
(A,A) (A _d ,A)	$\ell_0 = \frac{1}{m}(y_1 + \dots + y_m)$ $b_0 = \frac{1}{m} \left[\frac{y_{m+1}-y_1}{m} + \dots + \frac{y_{m+m}-y_m}{m} \right]$ $s_0 = y_m - \ell_0, s_{-1} = y_{m-1} - \ell_0, \dots, s_{-m+1} = y_1 - \ell_0$
(A,M) (A _d ,M)	$\ell_0 = \frac{1}{m}(y_1 + \dots + y_m)$ $b_0 = \frac{1}{m} \left[\frac{y_{m+1}-y_1}{m} + \dots + \frac{y_{m+m}-y_m}{m} \right]$ $s_0 = y_m/\ell_0, s_{-1} = y_{m-1}/\ell_0, \dots, s_{-m+1} = y_1/\ell_0$

A.4 - Formulas de cálculo recursivas

Trend	Seasonal		
	N	A	M
N	$\hat{y}_{t+h t} = \ell_t$ $\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$	$\hat{y}_{t+h t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1 - \gamma)s_{t-m}$
A	$\hat{y}_{t+h t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1 - \gamma)s_{t-m}$
A_d	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$
M	$\hat{y}_{t+h t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$\hat{y}_{t+h t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1 - \gamma)s_{t-m}$
M_d	$\hat{y}_{t+h t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$	$\hat{y}_{t+h t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^\phi) + (1 - \gamma)s_{t-m}$	$\hat{y}_{t+h t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^\phi)) + (1 - \gamma)s_{t-m}$

A.5.1 - Equações de espaço de estados para os modelos na estrutura ETS - Erros Aditivos

ADDITIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M_d	$y_t = \ell_{t-1}b_{t-1}^\phi + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$

A.5.2 - Equações de espaço de estados para os modelos na estrutura ETS - Erros Multiplicativos

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A_d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M_d	$y_t = \ell_{t-1}b_{t-1}^\phi(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

A.6 - Resultados obtidos com MsE, MdsE, MsAE e MdsAE

Método	Horizonte temporal						Média		Rank médio
	1	2	3	4	6	8	1-4	1-8	
MsE									
SESi	0.004	0.001	-0.002	0.000	-0.003	-0.002	0.001	-0.001	1.625
SBAi	0.001	-0.002	-0.004	-0.002	-0.004	-0.002	-0.002	-0.002	2.125
iMAPA	0.002	-0.002	-0.005	-0.004	-0.007	-0.007	-0.002	-0.004	3.813
FIXEDs	0.013	0.007	0.000	-0.002	-0.014	-0.022	0.005	-0.006	4.063
NAIVE	0.017	0.012	0.008	0.007	-0.001	-0.004	0.011	0.005	4.375
TSBi	0.012	0.009	0.007	0.008	0.006	0.007	0.009	0.008	4.750
CROSTONi	-0.013	-0.016	-0.018	-0.016	-0.017	-0.015	-0.016	-0.016	6.500
AUTOs	0.086	0.081	0.075	0.074	0.065	0.058	0.079	0.071	8.000
TSBs	0.113	0.108	0.102	0.100	0.091	0.083	0.106	0.097	9.000
SBAAs	0.152	0.148	0.144	0.144	0.138	0.135	0.147	0.142	10.125
CROSTONs	-0.145	-0.149	-0.154	-0.154	-0.161	-0.165	-0.151	-0.156	10.875
MdsE									
NAIVE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
SBAAs	-0.050	-0.051	-0.054	-0.054	-0.057	-0.059	-0.052	-0.055	2.000
FIXEDs	-0.052	-0.056	-0.058	-0.056	-0.061	-0.063	-0.056	-0.059	3.000
TSBs	-0.064	-0.067	-0.072	-0.074	-0.080	-0.087	-0.069	-0.076	4.000
AUTOs	-0.081	-0.085	-0.089	-0.094	-0.103	-0.106	-0.087	-0.095	5.000
CROSTONs	-0.162	-0.167	-0.171	-0.172	-0.184	-0.190	-0.168	-0.177	6.000
TSBi	-0.273	-0.276	-0.280	-0.281	-0.284	-0.287	-0.278	-0.281	7.125
SBAi	-0.271	-0.277	-0.282	-0.283	-0.288	-0.288	-0.278	-0.283	7.875
CROSTONi	-0.285	-0.289	-0.291	-0.291	-0.297	-0.299	-0.289	-0.293	9.000
SESi	-0.299	-0.301	-0.304	-0.306	-0.311	-0.313	-0.303	-0.307	10.188
iMAPA	-0.300	-0.302	-0.305	-0.307	-0.313	-0.313	-0.304	-0.307	10.813
MsAE									
TSBs	0.640	0.654	0.663	0.673	0.694	0.711	0.658	0.678	1.000
AUTOs	0.672	0.685	0.694	0.703	0.724	0.740	0.689	0.708	2.000
iMAPA	0.797	0.798	0.799	0.803	0.808	0.813	0.799	0.804	3.125
NAIVE	0.779	0.799	0.809	0.809	0.836	0.850	0.799	0.819	4.625
TSBi	0.818	0.820	0.821	0.825	0.829	0.834	0.821	0.826	4.875
SESi	0.821	0.822	0.823	0.827	0.831	0.835	0.823	0.828	5.875
SBAAs	0.811	0.818	0.823	0.829	0.841	0.852	0.820	0.832	6.500
SBAi	0.850	0.850	0.850	0.854	0.857	0.860	0.851	0.854	8.500
CROSTONs	0.830	0.839	0.845	0.853	0.870	0.883	0.842	0.858	9.000
CROSTONi	0.852	0.853	0.853	0.856	0.859	0.863	0.854	0.857	9.500
FIXEDs	0.905	0.915	0.923	0.933	0.952	0.969	0.919	0.938	11.000
MdsAE									
NAIVE	0.329	0.347	0.358	0.362	0.388	0.403	0.349	0.370	1.375
TSBs	0.342	0.354	0.365	0.371	0.385	0.398	0.358	0.374	1.625
AUTOs	0.369	0.378	0.386	0.396	0.414	0.425	0.382	0.399	3.000
FIXEDs	0.416	0.417	0.427	0.428	0.443	0.450	0.422	0.433	4.000
SBAAs	0.429	0.432	0.438	0.441	0.454	0.458	0.435	0.445	5.000
CROSTONs	0.441	0.451	0.460	0.461	0.480	0.490	0.453	0.468	6.000
TSBi	0.564	0.568	0.568	0.571	0.577	0.580	0.568	0.573	7.000
SBAi	0.570	0.573	0.575	0.576	0.581	0.587	0.574	0.578	8.375
CROSTONi	0.572	0.573	0.576	0.578	0.580	0.585	0.575	0.579	8.625
iMAPA	0.578	0.580	0.580	0.582	0.589	0.590	0.580	0.584	10.000
SESi	0.579	0.581	0.585	0.586	0.591	0.592	0.583	0.587	11.000

A.7 - Resultados obtidos com MsSE, MdsSE, MAsE e MdAsE

Método	Horizonte temporal						Média		Rank médio
	1	2	3	4	6	8	1-4	1-8	
MsSE									
iMAPA	3.569	3.440	3.370	3.417	3.447	3.353	3.449	3.419	1.000
SESi	3.643	3.514	3.444	3.492	3.521	3.427	3.523	3.493	2.125
TSBi	3.651	3.522	3.451	3.498	3.528	3.434	3.531	3.500	3.125
CROSTONi	3.728	3.596	3.526	3.571	3.599	3.502	3.605	3.572	4.250
SBAi	3.735	3.603	3.531	3.576	3.603	3.508	3.611	3.578	5.250
TSBs	3.625	3.667	3.655	3.762	4.070	4.761	3.677	3.955	5.500
AUTOs	3.661	3.708	3.693	3.799	4.110	4.801	3.715	3.994	6.750
NAIVE	5.312	5.285	5.236	5.191	5.280	5.246	5.256	5.248	8.000
SBA _s	75.440	75.937	76.454	77.108	78.527	80.337	76.235	77.604	9.000
CROSTON _s	86.200	86.784	87.394	88.141	89.759	91.809	87.130	88.701	10.000
FIXED _s	845.211	861.686	878.813	896.762	934.994	976.818	870.618	908.129	11.000
MdsSE									
NAIVE	0.109	0.121	0.129	0.131	0.151	0.162	0.123	0.138	1.375
TSBs	0.117	0.125	0.133	0.138	0.148	0.158	0.128	0.140	1.625
AUTOs	0.136	0.143	0.149	0.157	0.171	0.180	0.146	0.159	3.000
FIXED _s	0.173	0.174	0.183	0.184	0.196	0.202	0.179	0.188	4.000
SBA _s	0.184	0.187	0.192	0.194	0.206	0.210	0.189	0.198	5.000
CROSTON _s	0.194	0.203	0.212	0.212	0.231	0.240	0.205	0.219	6.000
TSBi	0.318	0.322	0.323	0.327	0.333	0.336	0.323	0.328	7.000
SBAi	0.325	0.329	0.331	0.331	0.337	0.345	0.329	0.335	8.438
CROSTONi	0.327	0.329	0.331	0.334	0.337	0.343	0.330	0.335	8.563
iMAPA	0.335	0.336	0.337	0.339	0.347	0.349	0.337	0.342	10.000
SESi	0.336	0.337	0.343	0.343	0.349	0.350	0.340	0.345	11.000
MASe									
TSBs	0.981	1.009	1.029	1.051	1.097	1.133	1.018	1.061	1.000
AUTOs	1.085	1.112	1.131	1.151	1.195	1.229	1.120	1.161	2.000
NAIVE	1.165	1.203	1.225	1.234	1.289	1.320	1.207	1.251	3.000
iMAPA	1.356	1.360	1.364	1.372	1.383	1.395	1.363	1.375	4.000
TSBi	1.420	1.423	1.426	1.434	1.444	1.454	1.426	1.436	5.563
SBA _s	1.397	1.411	1.422	1.435	1.462	1.485	1.416	1.442	6.125
SESi	1.428	1.431	1.434	1.442	1.452	1.462	1.434	1.444	6.750
CROSTON _s	1.403	1.423	1.438	1.454	1.491	1.519	1.430	1.464	7.563
SBAi	1.493	1.494	1.496	1.502	1.509	1.517	1.496	1.504	9.000
CROSTONi	1.499	1.500	1.502	1.508	1.515	1.523	1.502	1.510	10.000
FIXED _s	1.624	1.646	1.665	1.686	1.730	1.770	1.655	1.698	11.000
MdAsE									
TSBs	0.578	0.596	0.611	0.619	0.644	0.669	0.601	0.626	1.000
NAIVE	0.587	0.618	0.637	0.644	0.693	0.713	0.622	0.659	2.375
AUTOs	0.614	0.629	0.645	0.658	0.688	0.706	0.637	0.664	2.625
FIXED _s	0.685	0.696	0.708	0.720	0.737	0.756	0.702	0.722	4.000
SBA _s	0.703	0.714	0.725	0.738	0.757	0.774	0.720	0.741	5.000
CROSTON _s	0.743	0.759	0.771	0.781	0.813	0.831	0.764	0.790	6.000
TSBi	0.845	0.856	0.861	0.867	0.877	0.885	0.857	0.869	7.250
iMAPA	0.858	0.862	0.867	0.871	0.878	0.883	0.865	0.872	7.750
SBAi	0.860	0.865	0.869	0.874	0.883	0.891	0.867	0.877	9.063
SESi	0.864	0.869	0.872	0.877	0.885	0.893	0.871	0.879	10.000
CROSTONi	0.870	0.870	0.872	0.879	0.889	0.896	0.873	0.882	10.938

A.8 - Resultados obtidos com MsPIS e MsAPIS

Método	Horizonte temporal						Média		Rank médio
	1	2	3	4	6	8	1-4	1-8	
MsPIS									
iMAPA	-0.002	-0.005	-0.012	-0.024	-0.088	-0.228	-0.011	-0.069	1.438
SBAi	-0.001	-0.005	-0.015	-0.038	-0.153	-0.397	-0.015	-0.118	2.500
CROSTONi	0.013	0.036	0.065	0.096	0.125	0.073	0.053	0.080	3.188
SESi	-0.004	-0.013	-0.029	-0.057	-0.172	-0.397	-0.026	-0.130	3.563
NAIVE	-0.017	-0.046	-0.086	-0.125	-0.218	-0.319	-0.069	-0.153	5.000
TSBi	-0.012	-0.039	-0.081	-0.144	-0.357	-0.721	-0.069	-0.263	5.500
AUTOs	-0.086	-0.254	-0.501	-0.823	-1.697	-2.871	-0.416	-1.212	7.500
FIXEDs	-0.013	-0.106	-0.324	-0.697	-2.008	-4.163	-0.285	-1.442	7.563
TSBs	-0.113	-0.335	-0.661	-1.087	-2.239	-3.780	-0.549	-1.598	8.750
CROSTONs	0.145	0.414	0.793	1.273	2.490	4.018	0.656	1.774	10.000
SBA _s	-0.152	-0.477	-0.988	-1.698	-3.754	-6.702	-0.829	-2.687	11.000
MsAPIS									
TSBs	0.640	1.692	3.158	5.029	10.013	16.685	2.630	7.208	1.000
AUTOs	0.672	1.788	3.350	5.348	10.679	17.813	2.790	7.682	2.000
CROSTONs	0.830	2.221	4.148	6.582	12.949	21.280	3.445	9.299	3.625
iMAPA	0.797	2.165	4.086	6.539	13.024	21.578	3.397	9.344	4.125
TSBi	0.818	2.238	4.241	6.810	13.629	22.647	3.527	9.773	5.250
SESi	0.821	2.245	4.255	6.834	13.677	22.727	3.539	9.808	6.500
SBA _s	0.811	2.238	4.273	6.895	13.898	23.234	3.554	9.966	6.875
NAIVE	0.779	2.194	4.254	6.947	14.294	24.308	3.544	10.257	7.125
CROSTONi	0.852	2.341	4.447	7.152	14.337	23.838	3.698	10.277	9.125
SBAi	0.850	2.337	4.443	7.149	14.341	23.856	3.695	10.279	9.375
FIXEDs	0.905	2.463	4.637	7.391	14.585	23.981	3.849	10.460	11.000